

클래식 공연 활성화를 위한 효과적 가격 모델 수립



예술의전당-정형데이터 분석분야 어드밴스드 리그
Miraclassic de BOAZ

https://github.com/jwshin0908/BigContest_2023

신재욱 (jwshin0908@gmail.com)
김혜연 (sally879@naver.com)
박규연 (20203065@kookmin.ac.kr)
최유진 (youjin0450@naver.com)

목차

01 분석 개요

- 1.1 주제 소개
- 1.2 분석 배경 및 목적
- 1.3 분석 순서도

04 콘서트홀 좌석 그룹핑

- 4.1 모델링
- 4.2 좌석 그룹핑 결과

07 결론

- 7.1 활용방안 & 기대효과
- 7.2 의의 및 한계

02 데이터 수집 및 전처리

- 2.1 데이터 수집
- 2.2 데이터 전처리
- 2.3 파생변수

05 공연 수요 예측 모델링

- 5.1 모델링 & 튜닝
- 5.2 최적 모델 선정
- 5.3 변수 중요도 해석

08 툴·데이터 및 참고문헌

03 데이터 EDA

- 3.1 데이터 요약 및 시각화
- 3.2 좌석 시각화
- 3.3 가설 수립 및 검정

06 가격 모델링

- 6.1 가격 모델링 개요
- 6.2 가격 산출 공식
- 6.3 가격 산출 예시

Part 1

분석 개요

- 1.1 주제 소개
- 1.2 분석 배경 및 목적
- 1.3 분석 순서도



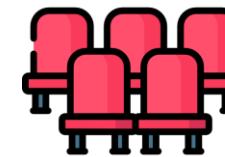
주제 소개



클래식 공연 활성화를 위한
예술의전당 음악당 콘서트홀의 효과적 가격 모델 수립



예술의전당 콘서트홀



좌석 그룹핑



공연별/좌석별 가격 모델

주제 소개

무료회원



유료회원



* 유니온페이 제휴회원 가입은 예술의전당 고객센터(1668-1352)에서 가능합니다.

가입한 무료 · 유료회원

클래식 공연에 대한
관람객의 관심도 수치화

관람객의 관심도에 따른
맞춤형 좌석 & 가격 제안

“ **멤버십을 고려한**
예술의전당 콘서트홀의 가격 모델 수립 ”

분석 배경

2023년 상반기 기준 공연시장에서의 클래식 공연

구분	공연건수 (건)	비중 (%)	공연회차 (회)	비중 (%)	티켓예매수 (건)	비중 (%)	티켓판매액 (원)	비중 (%)
전체	8,541	100	52,728	100	9,315,678	100	502,358,611,813	100
서양음악(클래식)	3,247 (▲12.1%)	38	3,640 (▲14.4%)	6.9	1,255,837 (▲38.3%)	13.5	35,435,840,680 (▲38.3%)	7.1

출처 : KOPIS. 2023년 상반기 공연시장 티켓판매 현황 분석 보고서

- 2023년 상반기 기준 전체 공연시장에서 클래식이 차지하는 비율은
공연건수 38%, 티켓판매수 13.5%, 티켓판매액 7.1%
- 가장 많은 공연이 이뤄진 장르는 서양음악(클래식) 38%,
그러나 티켓예매수와 티켓판매액은 상대적으로 티켓 가격이 높은 뮤지컬 장르에서 가장 높음 (41.9%, 45%)

관람객에게
합리적인 가격 제공

클래식 공연의 수요를 높이고
현재의 많은 공급을 충족할 수 있도록

예술의전당
수익 최대화

고객 관리의 중요성

고객과의 관계를 지속적으로 강화하기 위해서는

멤버십 가입 고객을 위한 맞춤형 관리 필요

고객 만족도 제고를 위해 멤버십을 고려한 가격 모델 수립 방안 제안

멤버십 고객 증가

회원제 사업 운영 아래
2022년 신규 유료 멤버십 증가율이
최대치를 기록하여
고객 관리의 필요성 대두

멤버십 고객의 중요성

신규 유료 멤버십의 증가는
기관 사업 활성화 및 예매율
상승에 기여하여
지속적인 관계 유지의 중요성 존재

분석 목적



다양한 관람객을 수용하는 가격 모델

- 현재의 가격 모델은 다양한 관람객의 관심과 선호도를 충분히 고려하지 못함
- 일부 관람객들은 원하는 공연에 접근하기에 어려움



- 관람객 특성과 공연의 성격을 반영한 새로운 좌석 그룹핑 기준 수립
- 관람객들에게 더 나은 즐거움 제공



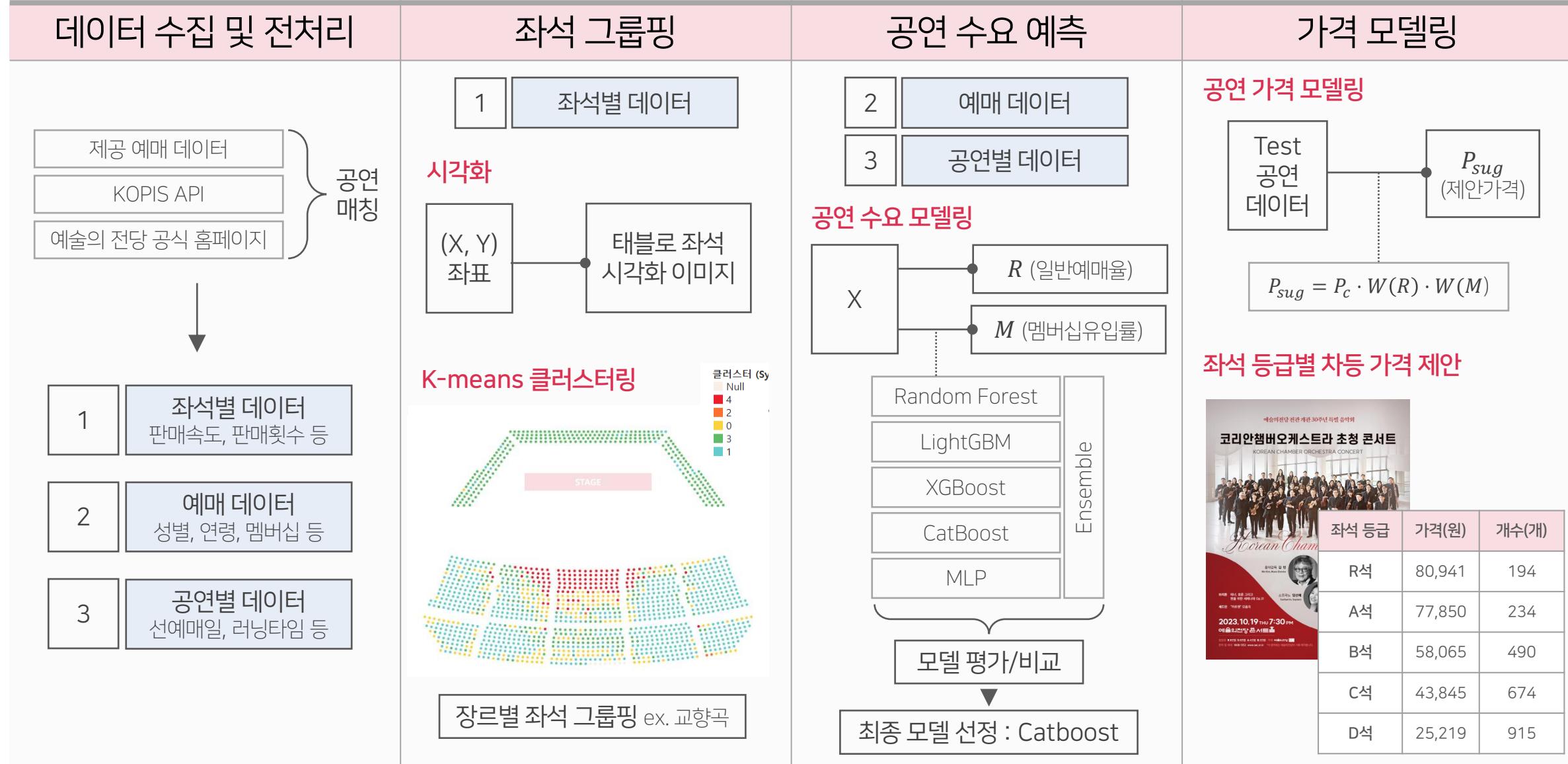
합리적인 가격 산정을 통한 투명성 제고

- 예매 과정에서 가격 산정에 대한 합리적 근거와 가격 정책의 투명성 부족



- 데이터 기반으로 합리적인 가격 모델 구축
- 가격 모델을 보다 투명하게 공개
- 관람객들에게 신뢰성 있는 정보 제공

분석 순서도



Part 2

데이터 수집 및 전처리

2.1 데이터 수집

2.2 데이터 전처리

2.3 파생변수



데이터 수집

제공 데이터

공연	공연 ID	공연일자	공연시간	예매시작일	선예매시작일	러닝타임
예매	공연 ID	좌석정보	할인내역	결제가격	예매일자·시간	
대관	공연 ID	공연일자	공연시간	장르		
회원	나이대	성별	멤버십	회원여부		

KOPIS API

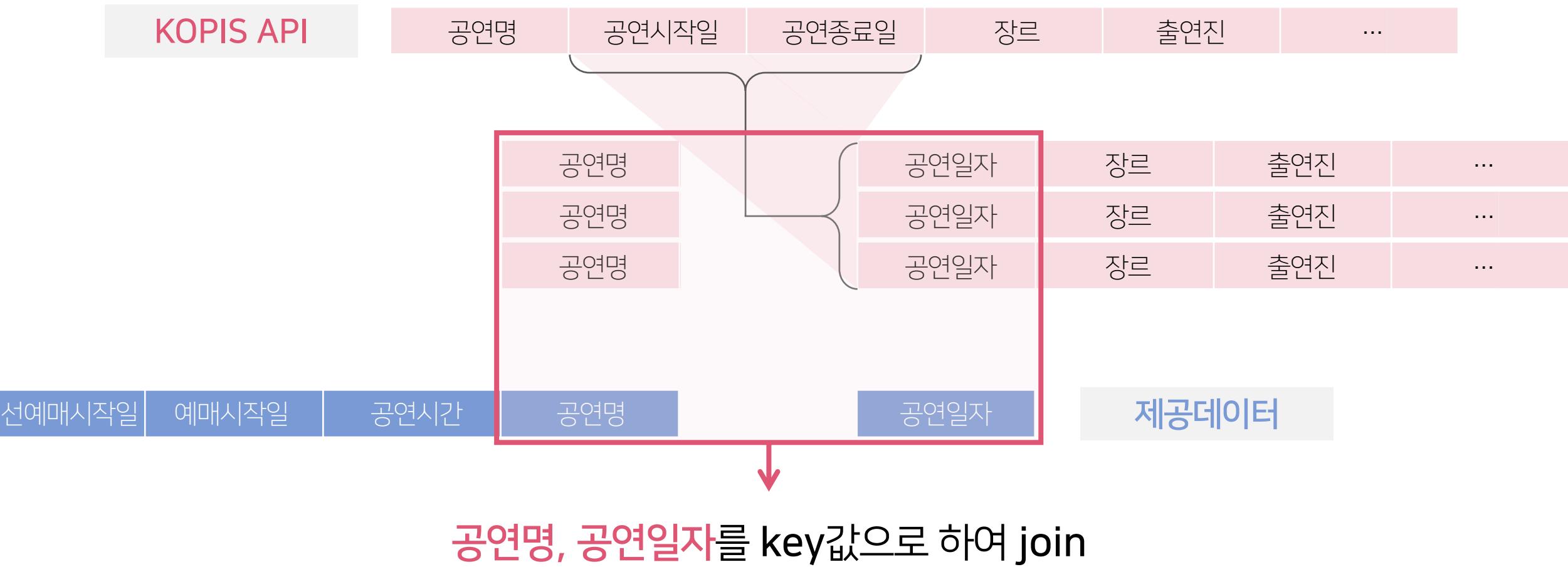
공연	공연명	공연시작일	공연종료일	장르	출연진	제작진
	관람연령	제작사	티켓가격	줄거리	공연상태	러닝타임
티켓 판매	공연일자	티켓판매수 (일)				

예술의전당 공식 홈페이지

공연 일정	공연명	장르	공연일자	장소
-------	-----	----	------	----

데이터 전처리

- (1) 제공 데이터와 추가 데이터 매칭



데이터 전처리

- (2) 제공 예매 데이터

컬럼	전처리 내용
공연 장소	<ul style="list-style-type: none"> 콘서트홀에 한정 (리사이틀홀, IBK챔버홀 공연 제거)
공연일자	예매 데이터의 가명정보처리로 인해 데이터 매칭 오차 발생
공연시작시간	<ul style="list-style-type: none"> 예술의전당 홈페이지 내 공연 세부 정보와 비교 → 실제로 공연하지 않은 데이터 제거
공연명 코드	<ul style="list-style-type: none"> 매칭되지 않는 공연 관련 컬럼 → 제공 데이터의 컬럼 값 사용
멤버십종류 1~6	<ul style="list-style-type: none"> 멤버십 종류별 컬럼을 생성 → {0, 1} bool 형태로 표기 무료회원 중 연령 관련 회원인 싹틔우미, 노블의 기준에 맞지 않는 회원 데이터는 제거
좌석번호	<ul style="list-style-type: none"> Split → 층, 구역, 열, 좌석번호 컬럼 생성 → 좌석에 대한 정보 구체적으로 표기

파생변수 - 관람객 관심도



멤버십에 따라 차등적으로 점수 정의
→ 클래식 공연에 대한 관람객의 관심도를 점수로 수치화

비회원 0점을 기준 → 유료회원의 연회비를 점수 간격에 활용

멤버십에 따른
관람객 관심도 파생변수 생성

멤버십종류	count	score
0 무료, 노블, 골드, 블루, 그린	82	25
1 무료, 싹티우미, 골드, 블루, 그린	778	25
2 무료, 골드, 블루, 그린	10097	23
3 무료, 싹티우미, 골드, 블루	1322	21
4 무료, 노블, 골드, 블루	1775	21
5 무료, 싹티우미, 골드, 그린	1297	19
6 무료, 골드, 블루	30009	19
7 무료, 노블, 골드, 그린	1624	19
8 무료, 골드, 그린	17967	17
9 무료, 싹티우미, 골드	2214	15
10 무료, 노블, 골드	2862	15
11 무료, 골드	69274	13
12 무료, 노블, 블루, 그린	244	13
13 무료, 싹티우미, 블루, 그린	1367	13
14 무료, 블루, 그린	14562	11
15 무료, 싹티우미, 블루	2911	9
16 무료, 노블, 블루	1094	9
17 무료, 노블, 그린	1999	7
18 무료, 싹티우미, 그린	10831	7
19 무료, 블루	25712	7
20 무료, 그린	48496	5
21 무료, 노블	3187	3
22 무료, 싹티우미	34787	3
23 무료	67162	1

파생변수 - 멤버십 등급

관람객 관심도를 등급으로 나누어 등급별로 선호하는 좌석에 차이가 있는지 확인하기 위함
 → 이후 EDA, 좌석 선호도 시각화에 활용



파생변수

- 공연 관련 파생변수

예매일자와 공연일자 간격

공연일자보다
얼마나 빨리 예매했는지를
나타내는 파생변수

공연 월

가정의 달 5월, 연말 12월 등
공연을 진행한 달을
나타내는 파생변수

선예매 여부

선예매기간이 존재할 경우 1,
아닌 경우 0

좌석 좌표

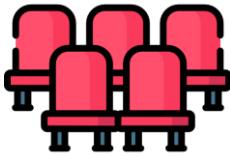
좌표평면 상에서의
좌석 위치에 대한 (x, y) 좌표

공연별 좌석 등급

공연별 가격 종류 카운트
→ 좌석 등급 개수 집계

파생변수

- 공연 관련 파생변수



오픈 좌석 수

코로나 거리두기 정책에 따라
제한적으로 오픈 된 좌석 수를
고려한 파생변수

~ 2020-06-07 : 2,505석
2020-06-28 ~ 2021-06-30 : 1,203석
2021-07-01 ~ : 2,505석



티켓 판매율

공연당 티켓 판매율을
나타내는 파생변수

실제 판매된 좌석 수
오픈된 좌석 수



멤버십 유입률

공연당 유료회원의 예매 정도를
나타내는 파생변수

실제 판매된 좌석이 갖는
관람객 관심도의 총합

최대 관람객 관심도 x 판매 좌석수

전처리 이후 최종 데이터

1

좌석별 데이터

	좌석	장르	총예매횟수	평균예매가격	평균예매와공연간격	평균멤버십점수	X	Y
0	1층 A블록10열 10	교향곡	75	31990.04975	366.63184	3.39801	-1.03268	-5.38793

3

공연별 데이터

	공연일자	공연시작시간	선예매여부	예매시작일	장르	휴게시간	공연명	내한여부	러닝타임	가격종류	R석	A석	요일	결제내역수	티켓판매량	전체좌석	일반예매율	멤버십합계	가중예매율	멤버십유입률
0	2018-11-25	1700	1	20181029.00000	클래식	15	김대진 피아노 협주곡	0	1시간 50분	3	50000.00000	20000.00000	6	1818	1537	2505.00000	0.61357	258	0.00412	0.00671

전처리 이후 최종 데이터

2

예매 데이터

연령	성별	예매일자	예매시간	공연일자	공연시작시간	좌석	결제가격	취소여부	할인내역	선예매시작일	예매시작일	장르	휴게시간	회원여부	공연명	내한여부	출연진	제작진	러닝타임	관람연령	제작사	가격종류	R석	S석	A석	티켓판매량	요일	멤버십	무료	싹티우미	노블골드	블루그린	증정	블록열	멤버십등급	좌석번호			
0	50.00000	M	20220206	1615	2022-03-02	1930	180000	0	일반	20220109.00000	20220110.00000	독주	0	Y	국립합창단위대한합창 I, 칼오르프&카르미나부라나	0	윤의중, 박미자, 박의준, 한명원 등	NaN	1시간10분	1만7세이상	NaN	[30000, 20000, 10000]	30000.00000	20000.00000	10000.00000	1039	2	무료	1	0	0	0	0	1	B	12	7	1	D

Part 3

데이터 EDA

3.1 데이터 요약 및 시각화

3.2 좌석 시각화

3.3 가설 수립 및 검정



데이터 요약 및 시각화 - (1) 전처리 이전 제공 데이터

기존 제공 데이터인 '2023빅콘테스트_어드밴스드리그_예술의전당.csv'를 활용해 특정 feature에 대해 시각화

데이터 형식 및 Null값 확인

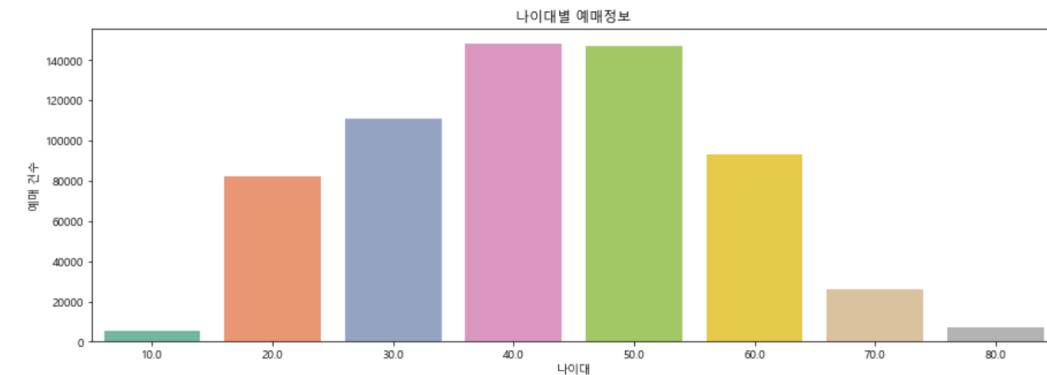
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1920868 entries, 0 to 1920867
Data columns (total 24 columns):
 #   Column           Dtype    
--- 
 0   age              float64  
 1   gender            object    
 2   membership_type_1 object    
 3   membership_type_2 object    
 4   membership_type_3 object    
 5   membership_type_4 object    
 6   membership_type_5 object    
 7   membership_type_6 float64  
 8   tran_date          int64    
 9   tran_time          int64    
 10  play_date          int64    
 11  play_st_time      int64    
 12  seat               object    
 13  price              int64    
 14  ticket_cancel     int64    
 15  discount_type     object    
 16  performance_code  int64    
 17  pre_open_date     float64  
 18  open_date          float64  
 19  genre              object    
 20  place              object    
 21  running_time       int64    
 22  intermission       int64    
 23  member_yn          object    
dtypes: float64(4), int64(9), object(11)
memory usage: 351.7+ MB
```

age	1300183
gender	1299714
membership_type_1	1299714
membership_type_2	1433215
membership_type_3	1749009
membership_type_4	1890675
membership_type_5	1919586
membership_type_6	1920868
tran_date	0
tran_time	0
play_date	0
play_st_time	0
seat	0
price	0
ticket_cancel	0
discount_type	0
performance_code	0
pre_open_date	864385
open_date	340
genre	25842
place	0
running_time	0
intermission	0
member_yn	0
dtype:	int64

EDA 과정을 거치며 수많은 Null값이
비회원 예매로부터 비롯되었음을 확인

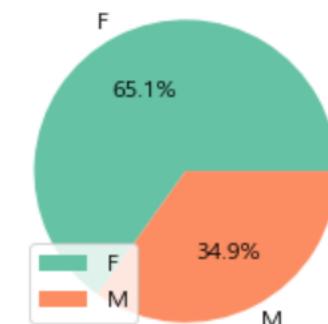
age(나이)

4, 50대의 예매 건수가 가장 높았음



gender(성별)

여성이 남성의 약 2배 정도의 비율을 차지

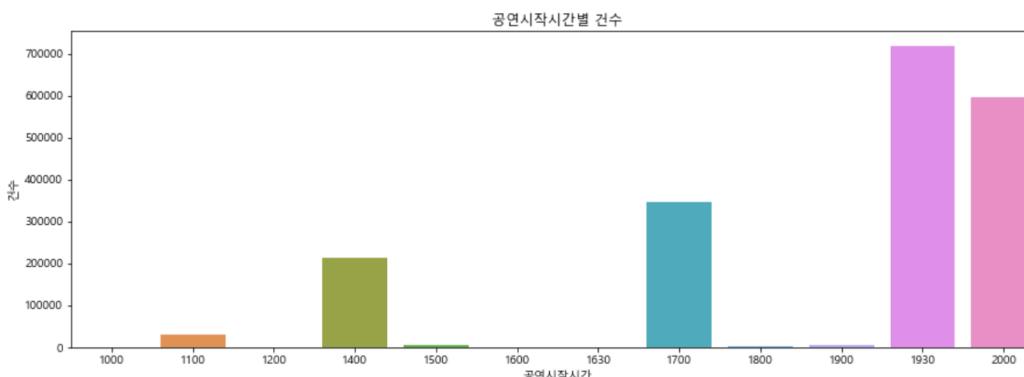


데이터 요약 및 시각화 - (1) 전처리 이전 제공 데이터

기존 제공 데이터인 '2023빅콘테스트_어드밴스드리그_예술의전당.csv'를 활용해 특정 feature에 대해 시각화

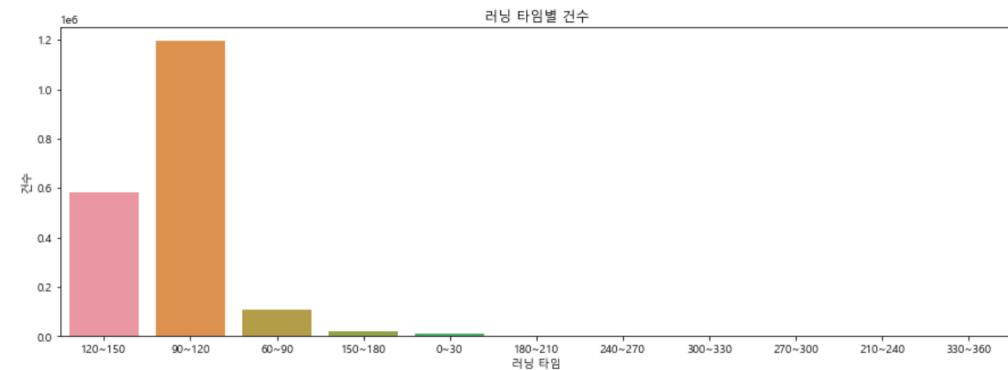
play_st_time(공연시작시간)

주로 저녁 시간이며, 19:30 이후의 공연이 큰 비율을 차지



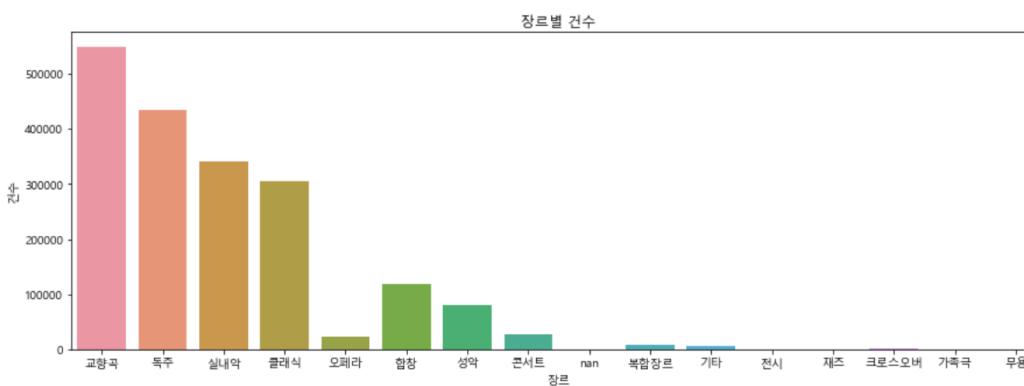
running_time(러닝타임)

90분에서 120분 정도의 러닝타임 공연 예매 건수가 가장 많았음



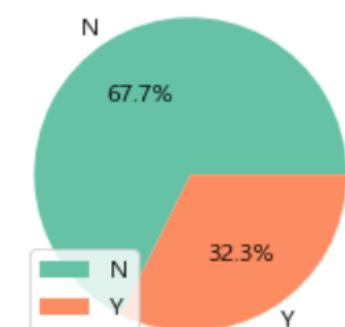
genre(장르)

교향곡 - 독주 - 실내악 순서로 예매 건수가 많음



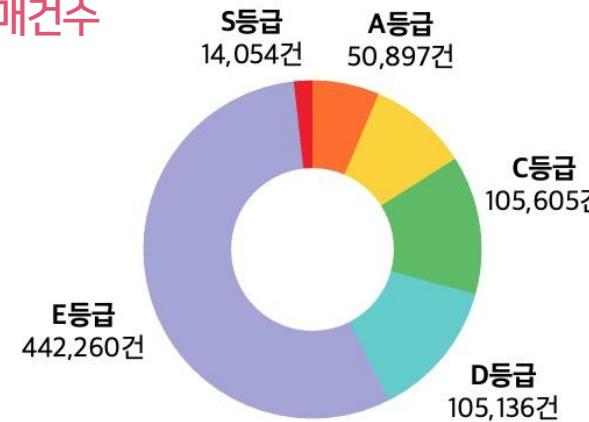
member_yn(회원 여부)

비회원이 회원보다 예매 건수를 2배 이상 차지



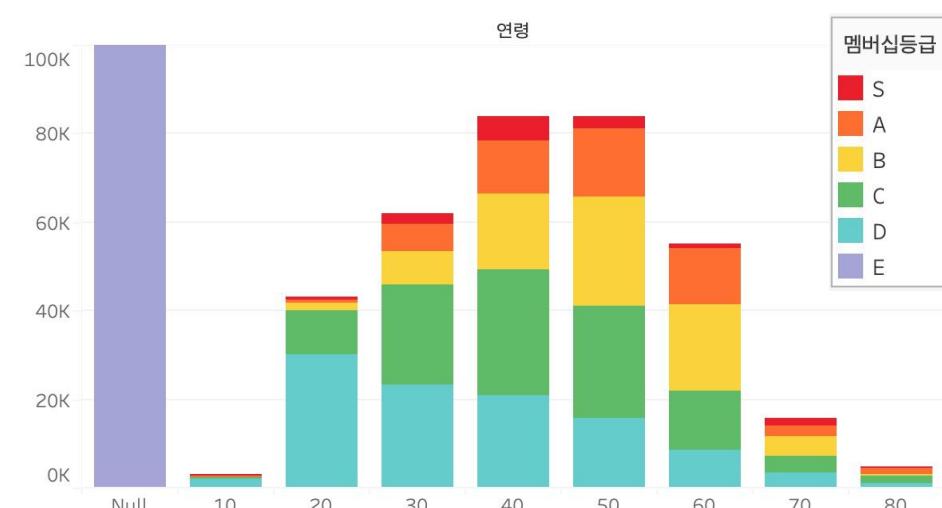
데이터 요약 및 시각화 - (2) 전처리 이후 예매 데이터 시각화 : 멤버십을 중심으로

멤버십 등급별 예매건수



연령별 멤버십 등급 분포

연령이 올라갈수록 높은 멤버십 등급을 보유



멤버십 등급별 선호 장르

모든 멤버십 등급이 교향곡, 클래식을 가장 많이 예매

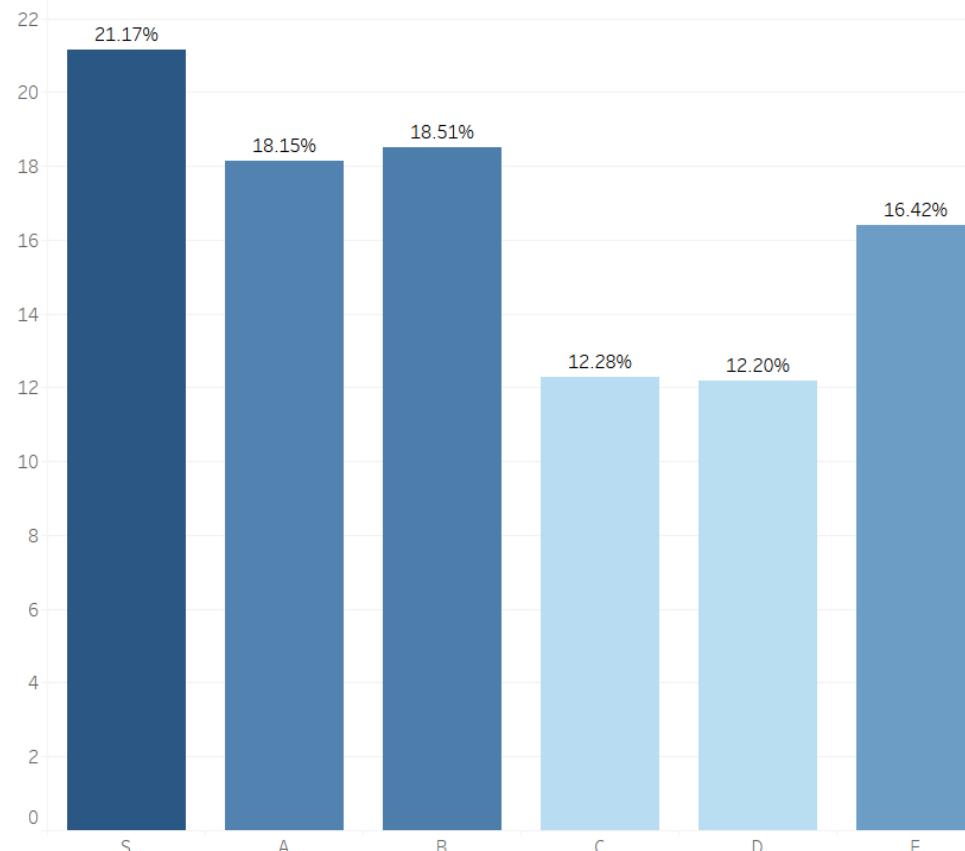
높은 멤버십 등급은 독주를,
낮은 멤버십 등급은 합창과 성악, 오페라를 선호하는 경향을 보임

멤버십등급	교향곡	기타	장르									
			독주	복합장르	성악	실내악	오페라	재즈	콘서트	크로스오..	클래식	합창
S	58.89%		13.04%	0.20%	1.59%	2.89%	0.50%	0.16%	0.18%	0.05%	20.31%	2.20%
A	59.75%	0.02%	11.57%	0.40%	2.18%	3.03%	0.73%	0.18%	0.37%	0.03%	19.15%	2.59%
B	57.27%	0.01%	11.67%	0.65%	2.52%	2.90%	0.89%	0.23%	0.53%	0.05%	19.99%	3.30%
C	52.89%	0.01%	14.58%	1.07%	2.80%	2.94%	0.91%	0.14%	1.20%	0.07%	20.62%	2.77%
D	54.12%	0.11%	8.01%	1.59%	2.16%	3.57%	0.94%	0.19%	0.65%	0.05%	21.86%	6.74%
E	46.55%	1.29%	1.65%	0.84%	4.31%	2.89%	2.93%	0.10%	2.47%	0.17%	24.01%	12.78%

데이터 요약 및 시각화 - (2) 전처리 이후 예매 데이터 시각화 : 멤버십을 중심으로

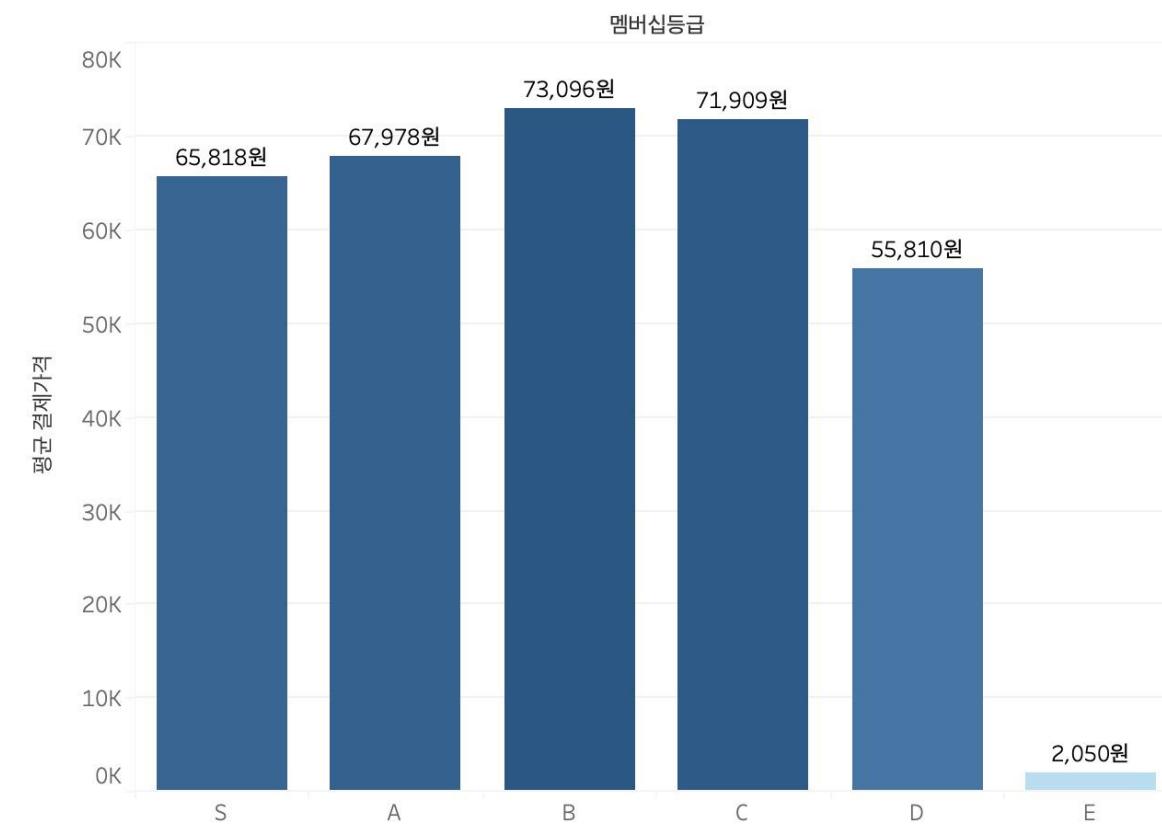
멤버십 등급별 평균 할인율

S등급이 평균적으로 가장 높은 할인율로 티켓을 예매하였고, 이어서 골드를 포함한 유료회원을 보유한 A, B등급의 평균 할인율이 높음 비회원인 E등급의 대부분이 초대권 할인을 받아 높은 평균 할인율을 보임



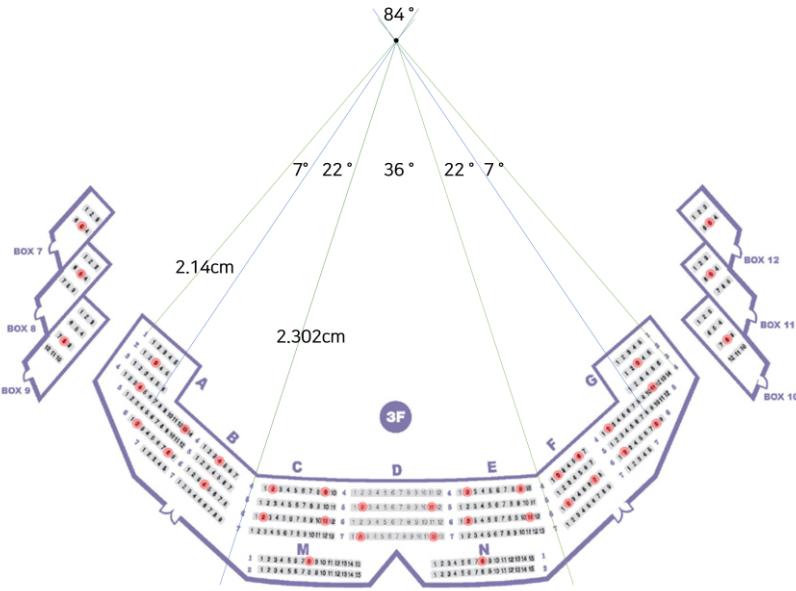
멤버십 등급별 평균 결제가격

S, A 등급이 정가가 높은 티켓을 예매하지만 보유한 멤버십에 따라 큰 폭으로 할인을 받으며 B, C 등급보다 평균 결제가격이 낮음 비회원인 E등급의 대부분이 초대권 할인을 받아 평균 가격이 매우 낮음



좌석 시각화 - (1) 태블로를 활용한 좌석 배치도 구현

홈페이지 좌석 배치도상 각도 · 간격 계산



홈페이지 좌석 배치도 이미지의
1, 2, 3층을 구분하여
가상의 원점을 설정하고,
원점에서부터 각 좌석의 간격과 각도 계산

좌표 평면에 좌석을 점으로 구현

```
# 맨 처음 부채꼴 중심 좌표와 반지를 정부
center_x = 0
center_y = 0
radius = 2.14

# 중심각과 시작 각도
central_angle = 22
start_angle = 230
end_angle = start_angle + central_angle

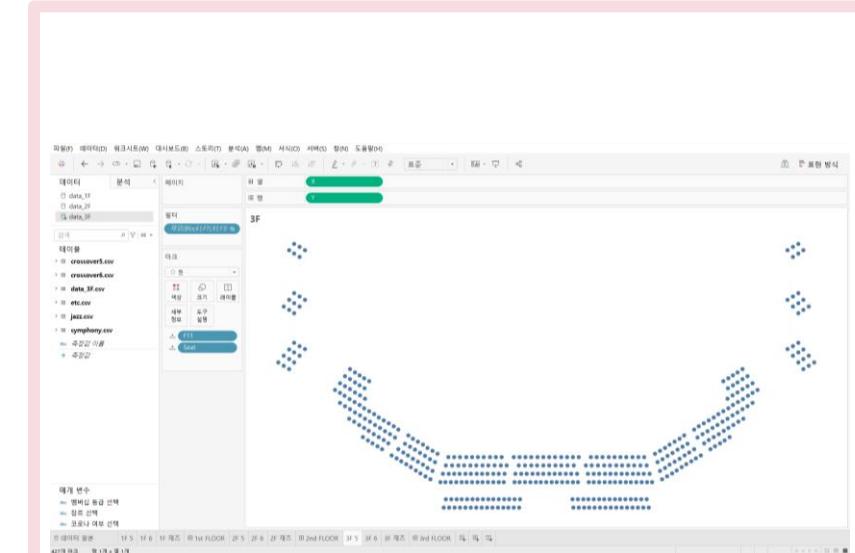
# 부채꼴의 양 끝 점 계산
start_point_x = center_x + radius * 
np.cos(np.radians(start_angle))
start_point_y = center_y + radius * 
np.sin(np.radians(start_angle))

end_point_x = center_x + radius * 
np.cos(np.radians(end_angle))
end_point_y = center_y + radius * 
np.sin(np.radians(end_angle))

# 현 위의 점 좌표 계산
나머지 행 좌표 계산 반복 및 데이터프레임 저장
```

원점과 좌석의 간격을 반지름으로 설정하고,
부채꼴의 현 또는 호 위에
일정한 간격으로 점을 찍는 방식으로
좌석 좌표 계산

태블로 좌석 시각화



생성한 x, y 좌표 데이터를 활용하여
태블로로 좌석을 좌표평면에 원으로 시각화
이후 원하는 기준에 따라 색상에 차이를 두어
각 좌석의 특성을 시각화

좌석 시각화 - (2) 관람객 관심도에 따른 좌석 선호도

장르 및 멤버십 등급별로 예매 데이터를 합하여
장르별로 각 멤버십 등급이 좌석을 얼마나 예매하였는지 확인

1F S 등급의 교향곡 총 예매 좌석



1F A 등급의 교향곡 총 예매 좌석

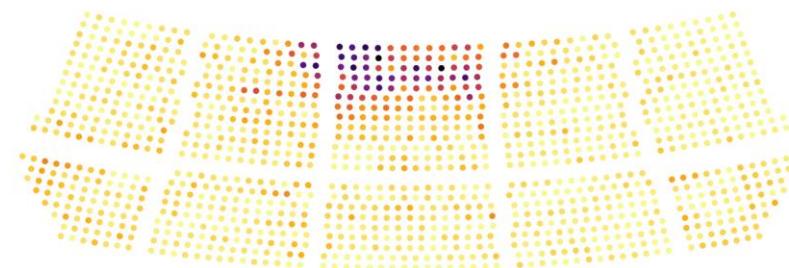


1F E 등급의 교향곡 총 예매 좌석



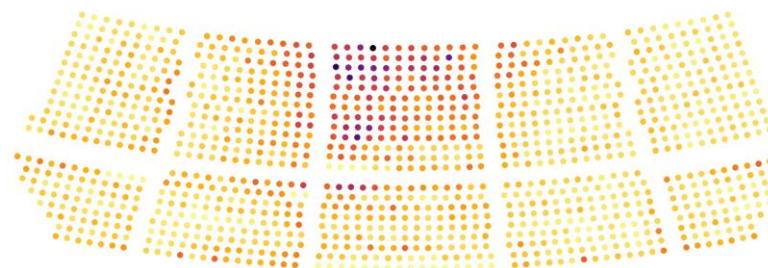
S 등급의 교향곡 예매 현황

앞쪽, 왼쪽 좌석 선호



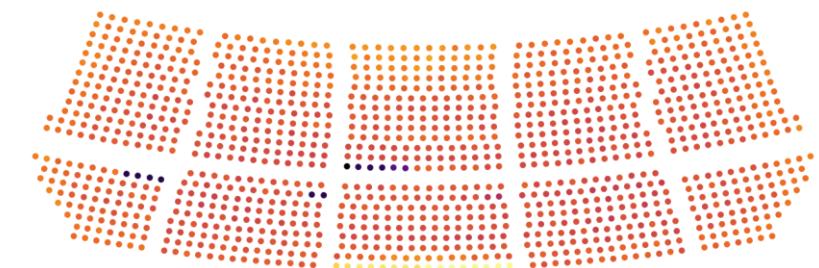
A 등급의 교향곡 예매 현황

앞쪽, 왼쪽 좌석 선호



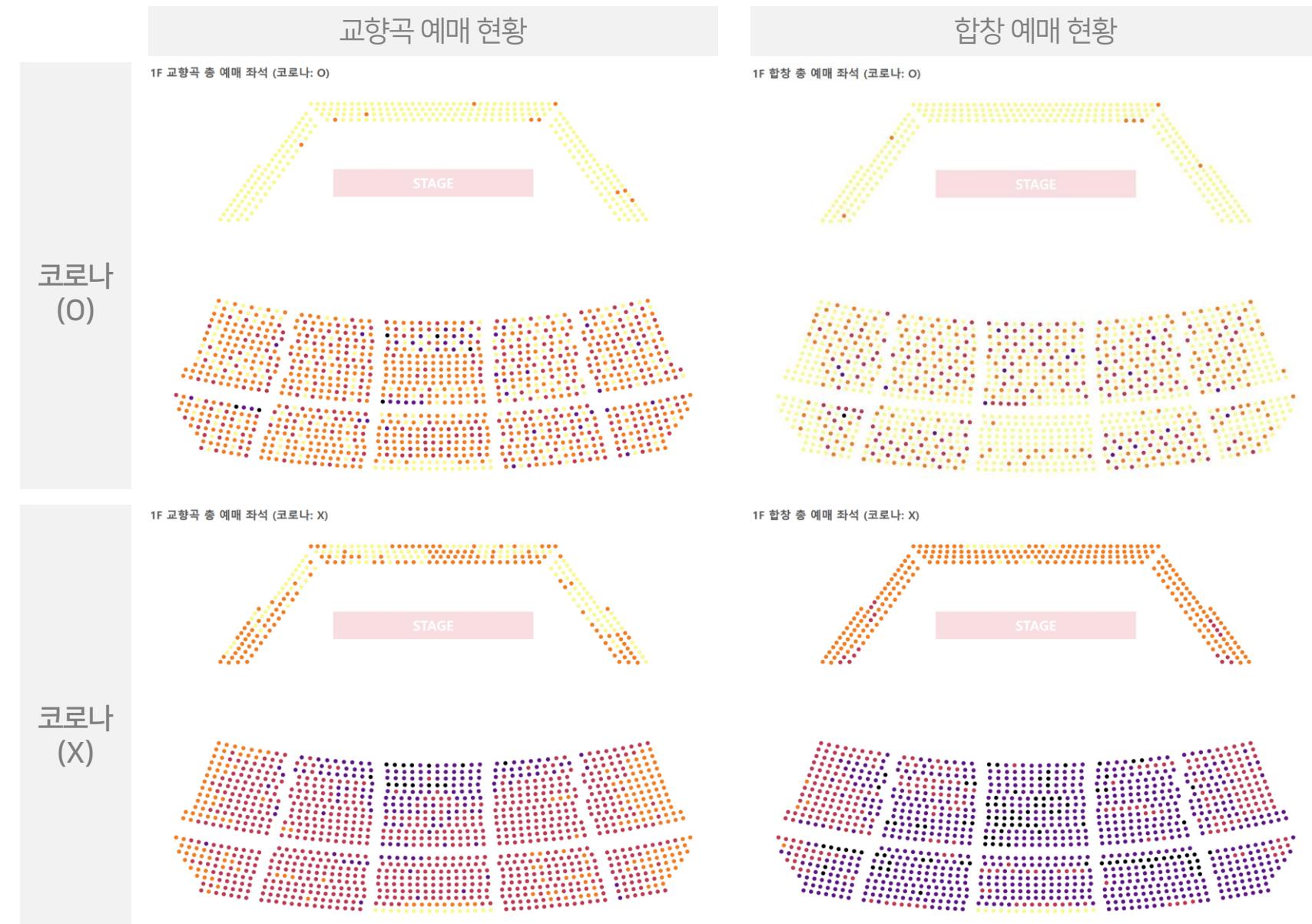
E 등급의 교향곡 예매 현황

중간, 뒤쪽 좌석 선호



좌석 시각화 - (3) 장르 및 코로나 여부에 따른 좌석 예매 현황

예매 데이터 중
코로나의 영향을 고려해야 하는
2020년, 2021년의 데이터,
나머지 기간의 데이터를
구분하여
코로나 여부와 장르별로 좌석이
얼마나 예매되었는지 확인



가설 수립 및 검정

가설

멤버십
가입에
따라

공연에 대한 지불 용의가 다를 것이다.

선호하는 장르가 다를 것이다.

선호하는 좌석이 다를 것이다.

이용하는 할인 내역이 다를 것이다.

장르에 따라

선호하는 좌석이 다를 것이다.

일찍 예매된 좌석일수록

좋은 좌석일 것이다.

선예매 공연은

인기가 많은 공연일 것이다.

검정 및 활용

독립성 검정 결과 $p\text{-value} < 0.05$ 로
관람객 등급과 결제 가격, 선호 장르가
관련이 있다고 결론

관람객 관심도에 따른 좌석과 할인 내역 시각화
→ 멤버십 가입에 따라 다른 경향성을 보임

다양한 리뷰 및 컬럼, 선행연구에서 확인
→ 장르별 좌석 그룹핑 진행

가정하고 예매일자와 공연일자 사이의 간격을
좌석 그룹핑 모델에서 변수로 활용

직관적인 사실로, 선예매 여부를
모델 내의 독립변수로 활용

Part 4

콘서트홀 좌석 그룹핑

4.1 모델링

4.2 좌석 그룹핑 결과



모델링 - (1) 좌석별 데이터 생성

클러스터를 좌석 좌표에 구현하기 위해

예매 데이터를 **개별 좌석을 기준으로 groupby**하여 좌석별 데이터 생성

예매 데이터

컬럼 생성

총 예매 횟수가 **많은**

평균 예매 가격이 **높은**

평균 예매일자 공연일자 간격이 **긴**

평균 멤버십 점수가 **높은**

태블로 좌석 구현

좌표 평면상 X 좌표

좌표 평면상 Y 좌표

좌석이 **인기가 많은 좌석**이라고 정의

좌석별 데이터 + 위치 데이터

최종 좌석별 데이터

좌석	장르	총예매횟수	평균예매가격	평균예매와공연간격	평균멤버십점수	X	Y
0 1층 A블록10열 10	교향곡	75	31990.04975	366.63184	3.39801	-1.03268	-5.38793
1 1층 A블록10열 10	기타	0	0.00000	60.00000	0.00000	-1.03268	-5.38793
2 1층 A블록10열 10	독주	15	73952.38095	969.71429	6.23810	-1.03268	-5.38793
3 1층 A블록10열 10	복합장르	2	22500.00000	2304.25000	7.00000	-1.03268	-5.38793
4 1층 A블록10열 10	성악	8	37500.00000	99.85000	3.70000	-1.03268	-5.38793
5 1층 A블록10열 10	실내악	8	51533.33333	178.66667	4.40000	-1.03268	-5.38793
6 1층 A블록10열 10	오페라	1	3000.00000	69.14286	1.85714	-1.03268	-5.38793

모든 변수를 고려하고자

모든 데이터 MinMax Scaling 후

클러스터링 진행

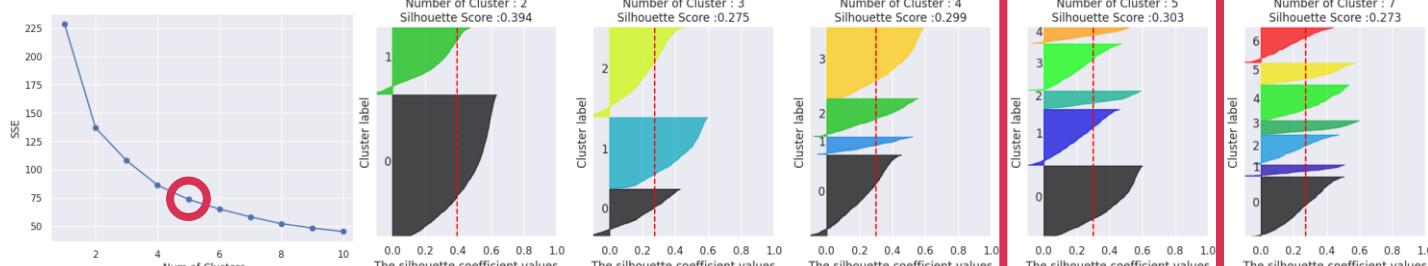
모델링 - (2) 클러스터링

장르별로 선호되는 좌석이 다를 것이라고 예상하여

장르별로 데이터를 나눈 좌석별 데이터로 k-means 클러스터링 진행

엘보우 그래프와 실루엣 계수의 평균값, 분포 → 최적 클러스터링 개수 k 선정
4개 컬럼별 센터값의 순위가 높은 클러스터 → 인기가 많은 클러스터로 설정

ex) 교향곡 좌석별 데이터 클러스터링 과정

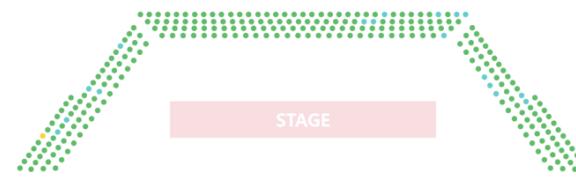


k = 5

클러스터별로 4개 컬럼별 센터값의 순위 도출
순위의 평균값을 기준으로 오름차순 정렬
→ 상위에 위치할 수록 인기 많은 클러스터

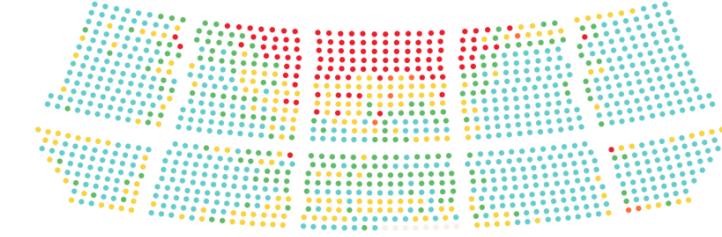
— 클러스터링 결과 →

교향곡 좌석 클러스터링



STAGE

교향곡 좌석 클러스터링



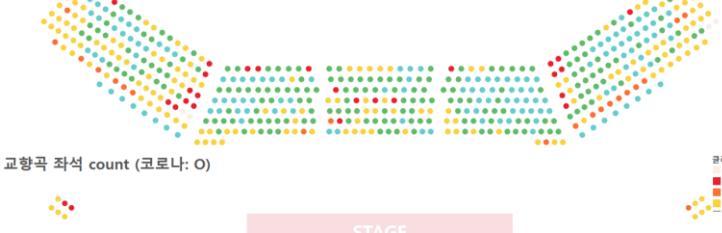
STAGE

교향곡 좌석 count (코로나: O)



STAGE

교향곡 좌석 count (코로나: X)



STAGE

좌석 그룹핑 결과

클러스터링을 통한 장르별 좌석 그룹핑 결과

교향곡 k=5

- | | |
|------------|------------|
| R: 4(194개) | A: 2(221개) |
| B: 0(560개) | C: 3(673개) |
| D: 1(861개) | |

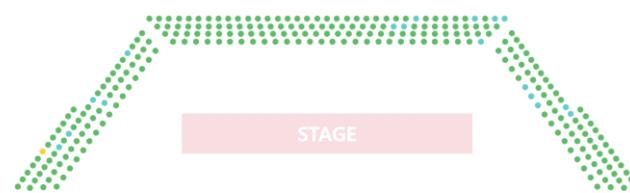
클래식 k=5

- | | |
|------------|------------|
| R: 2(194개) | A: 4(234개) |
| B: 0(490개) | C: 3(674개) |
| D: 1(915개) | |

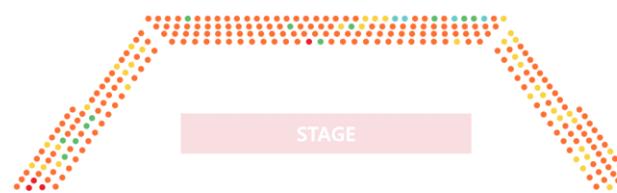
합창 k=4

- | | |
|------------|------------|
| R: 2(396개) | A: 1(835개) |
| B: 3(410개) | C: 0(865개) |

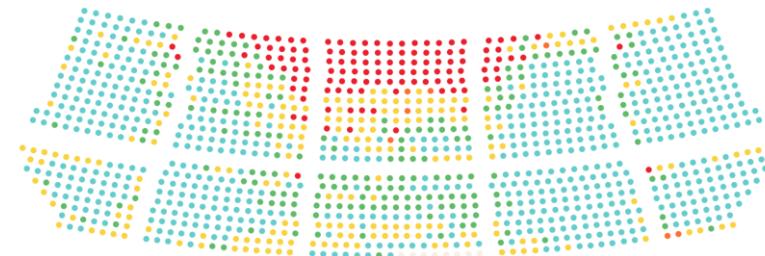
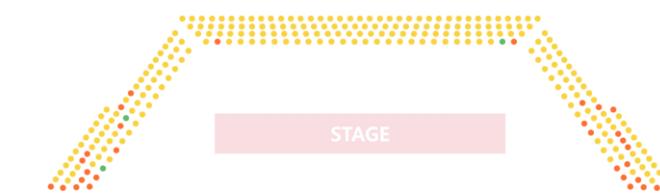
교향곡 좌석 클러스터링



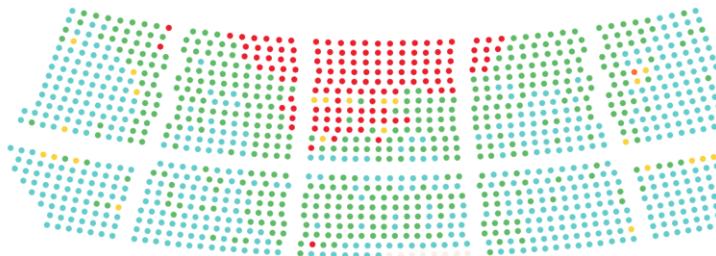
클래식 좌석 클러스터링
클러스터 (S)



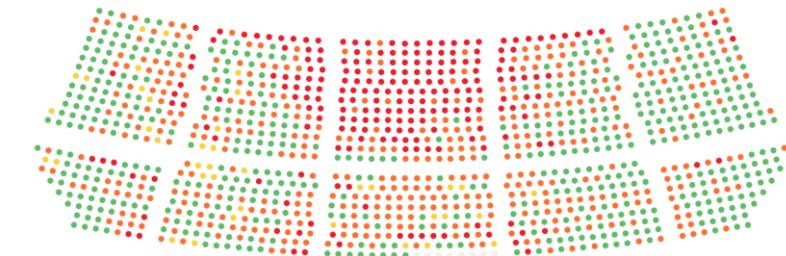
합창 좌석 클러스터링



D블록 < B블록(왼쪽) 선호
협연자가 오케스트라를 등지고
지휘자의 왼쪽에 위치



교향곡에 비해 중앙 부분 선호
한쪽으로 치우쳐 앉으면
특정 악기의 소리만 강하게 들릴 수 있음



중앙 부분 넓은 구역 선호
다양한 음성의 소리를
비슷한 거리에서 감상할 수 있음

Part 5

공연 수요 예측 모델링

- 5.1 모델링 & 튜닝
- 5.2 최적 모델 선정
- 5.3 변수 중요도 해석



모델링 & 튜닝 - (1) X, y 변수

R (일반예매율), M (멤버십 유입률)을 예측하는 회귀모델 학습을 위해 공연별 데이터를 변형해 일부 feature 사용

- X : 7개 변수
- Y : R (일반예매율), M (멤버십 유입률)

X										Y
공연일자	공연 시작시간	선예매여부	예매시작일	장르	휴게시간	공연명	내한여부	러닝타임	가격종류	R (일반예매율)
R석	A석	요일	결제내역수	티켓판매량	전체좌석	멤버십합계	가중예매율	월		M (멤버십 유입률)

$$R = \text{일반예매율} = \frac{n}{N}$$

N = 오픈 좌석 수

$$R_w = \text{멤버십가중예매율} = \frac{\sum_{i=1}^n m_i}{25N}$$

n = 판매 좌석 수

$$M = \text{멤버십 유입률} = \frac{R_w}{R} = \frac{\sum_{i=1}^n m_i}{25n}$$

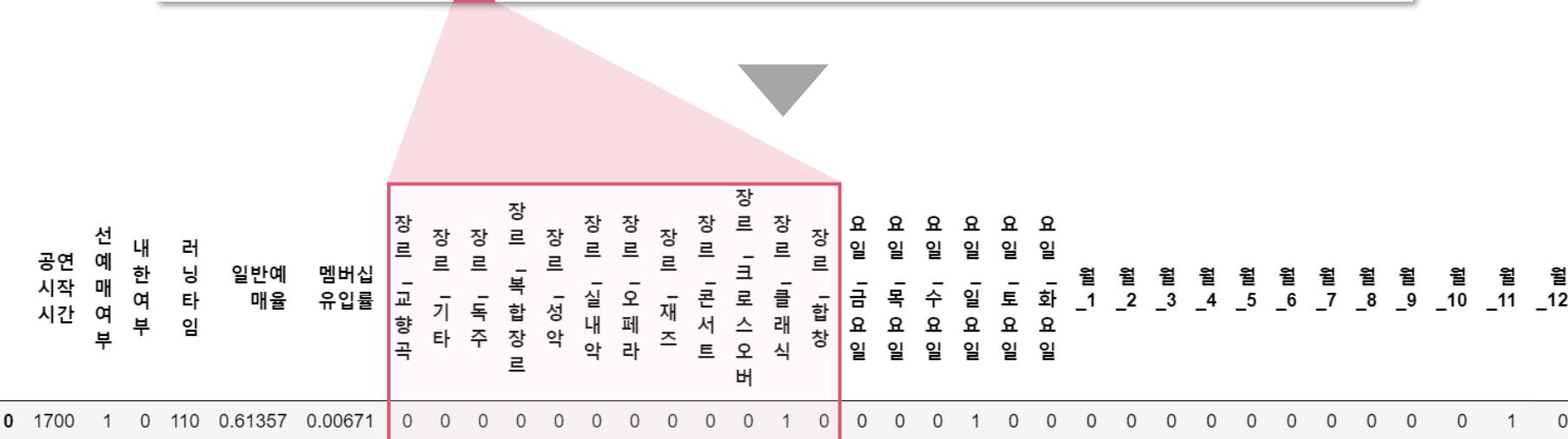
$m_i = i$ 번째 예매 데이터의 멤버십 점수

모델링 & 튜닝 - (2) 범주형 데이터

범주형 데이터 더미화(one-hot encoding)

3개 범주형 컬럼(장르, 요일, 월) → 30개 컬럼

공연 일자	공연 시작 시간	선 예매 여부	예매시작일	장르	휴게 시간	공연명	내한여부	러닝타임	가격 종류	R석	A석	요일	결제 내역 수	티켓 판매량	전체좌석	일반예매율	멤버십 할계	가중예매율	멤버십 유입률	월
0 2018-11-25	1700	1	20181029.000000	클래식	15	김대진 피아노 독주회	0	110	3	50000.000000	20000.000000	일요일	1818	1537	2505.000000	0.61357	258	0.00412	0.00671	11



모델링 & 튜닝 - (3) train/test 데이터 분리

하나의 데이터셋을 train/test 데이터셋으로 분리



이후 Grid Search 진행 시 Cross Validation (교차 검증) 진행, CV=5



모델링 & 튜닝 - (4) 연속형 데이터

연속형 데이터 Standard Scaling

Train 데이터셋 scaling한 만큼 Test 데이터셋 scaling 진행

공연시작시간	러닝타임
329	1930
173	1930
272	1930
496	1930
182	1930
	110
	120
	120
	120
	90



- Train 데이터셋

공연시작시간	러닝타임
0	0.42978
1	0.42978
2	0.42978
3	0.42978
4	0.42978
	-0.14218
	0.55189
	0.55189
	0.55189
	-1.53034

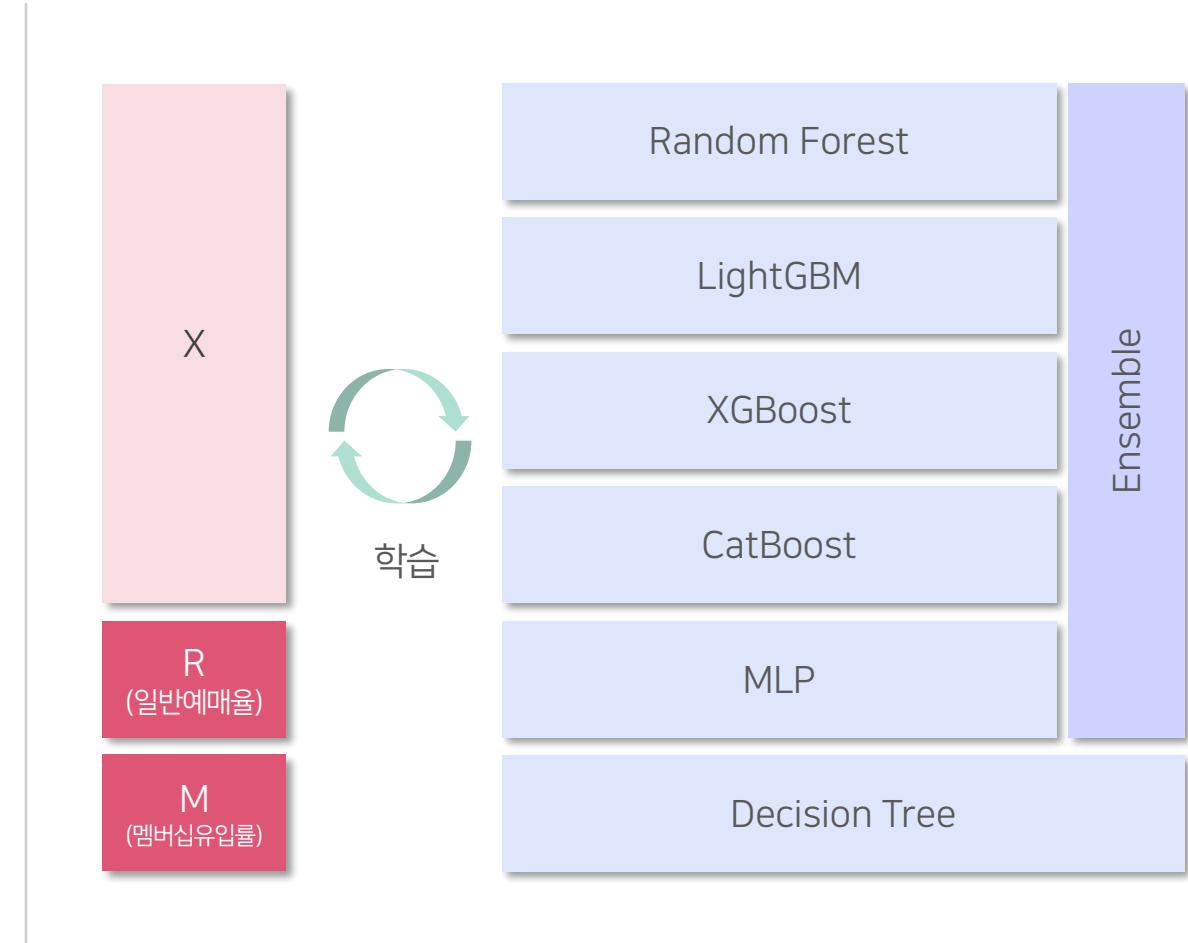
- Test 데이터셋

공연시작시간	러닝타임
0	0.42978
1	0.42978
2	-1.08252
3	0.42978
4	-1.08252
	-0.83626
	0.55189
	-0.14218
	-0.14218
	0.55189

모델링 & 튜닝 - (5) 모델 학습

R (일반예매율), M (멤버십 유입률) 예측을 위한 각 예측 y 별 회귀모델 선택 및 학습 진행

모델	설명
Decision Tree	데이터 분류와 예측을 위한 트리 구조의 알고리즘으로, 데이터의 특성에 따라 의사결정 규칙을 생성하여 문제를 해결함
Random Forest	여러 개의 의사결정나무(Decision Tree)를 조합하여 안정적이고 정확한 예측을 수행하는 앙상블(Ensemble) 기법
LightGBM	일반적인 그래디언트 부스팅(Gradient Boosting) 방법을 사용하면서도 빠른 학습과 예측 속도, 과적합 방지 등의 성능 향상을 위한 최적화 기법을 적용한 부스팅 알고리즘
XGBoost	트리 분류기와 부스팅 알고리즘을 활용하여 높은 예측 성능을 가지며, 과적합 방지와 빠른 학습 및 예측 속도 등이 특징인 알고리즘
CatBoost	범주형 특성을 다루는데 있어서 탁월한 예측 성능을 가지며, 자체적으로 과적합을 방지하는 기능을 가진 부스팅 알고리즘
MLP	인공 신경망 기반의 다층 퍼셉트론으로, 복잡한 비선형 문제를 다룰 수 있는 대표적인 머신러닝 알고리즘
Ensemble	여러 개의 머신러닝 모델을 결합하여 예측을 수행하는 앙상블 기법 본 실험에서는 성능 상위 5개의 머신러닝 모델들을 결합함(Decision Tree 제외)



모델링 & 튜닝 - (6) 모델 하이퍼파라미터 튜닝

Grid Search CV를 진행하여 최적의 하이퍼파라미터 선정

Decision Tree	Random Forest	LightGBM
<pre>param_grid = { 'max_depth': [None, 10, 20, 30], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4] }</pre>	<pre>param_grid = { 'n_estimators': [50, 100], 'max_depth': [5, 10], 'min_samples_split': [5, 10], 'min_samples_leaf': [1, 2, 4] }</pre>	<pre>param_grid = { 'num_leaves': [5, 10], 'learning_rate': [0.001, 0.01, 0.1], 'n_estimators': [100, 500, 1000] }</pre>
XGBoost	CatBoost	MLP
<pre>param_grid = { 'n_estimators': [50, 100, 200], 'max_depth': [3, 5, 7], 'learning_rate': [0.001, 0.01, 0.1] }</pre>	<pre>param_grid = { 'depth': [3, 5, 7, 10, 20], 'learning_rate': [0.001, 0.01, 0.1, 0.3, 0.5] }</pre>	<pre>param_grid = { 'hidden_layer_sizes': [(10,), (50,)], 'activation': ['logistic', 'tanh', 'relu'], 'solver': ['lbfgs', 'sgd', 'adam'], 'alpha': [0.001, 0.01, 0.1] }</pre>

최적 모델 선정 - (1) R (일반예매율) 예측

모델 비교

평가 지표 \ 모델	Decision Tree	Random Forest	LightGBM	XGBoost	CatBoost	MLP Regressor	Ensemble (상위 5개 모델 앙상블)
MSE	0.0945	0.0730	0.0717	0.0726	0.0716	0.0721	0.0716
MSLE	0.0413	0.0322	0.0317	0.0323	0.0317	0.0324	-

최적 모델 선정

CatBoost의 MSE가 가장 낮으므로,
일반예매율 예측 모델로 CatBoost 선정

최적 모델 선정 - (2) M (멤버십 유입률) 예측

모델 비교

평가 지표 \ 모델	Decision Tree	Random Forest	LightGBM	XGBoost	CatBoost	MLP Regressor	Ensemble (상위 5개 모델 앙상블)
MSE	0.0157	0.0146	0.0146	0.0145	0.0144	0.0130	0.0140
MSLE	0.0117	0.0110	0.0109	0.0109	0.0985	0.0097	0.0105

최적 모델 선정

MLP의 MSE가 가장 낮지만 모델의 해석력을 위해
 멤버십 유입률 예측 모델로 CatBoost 선정

변수 중요도 해석 - (1) R(일반예매율) 예측 : SHAP Feature Importance

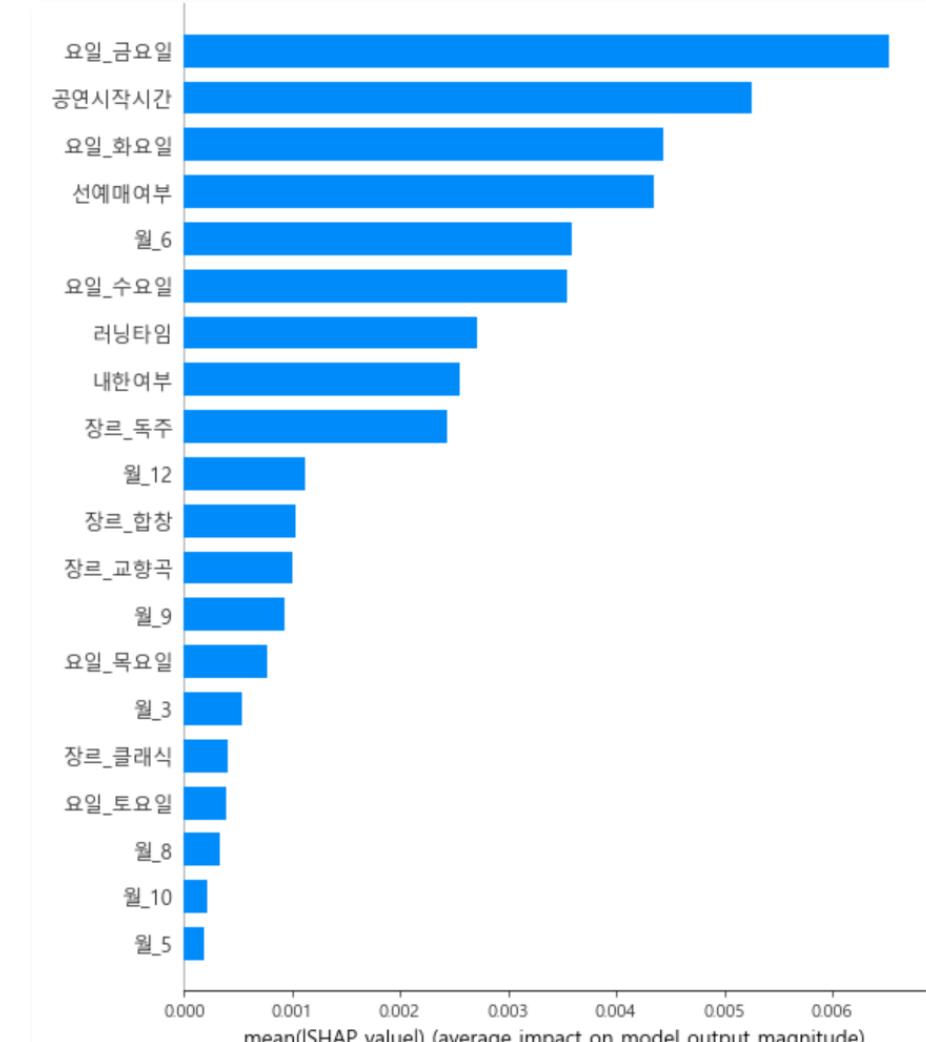
SHAP 변수 중요도

Shap (SHapley Additive exPlanations)

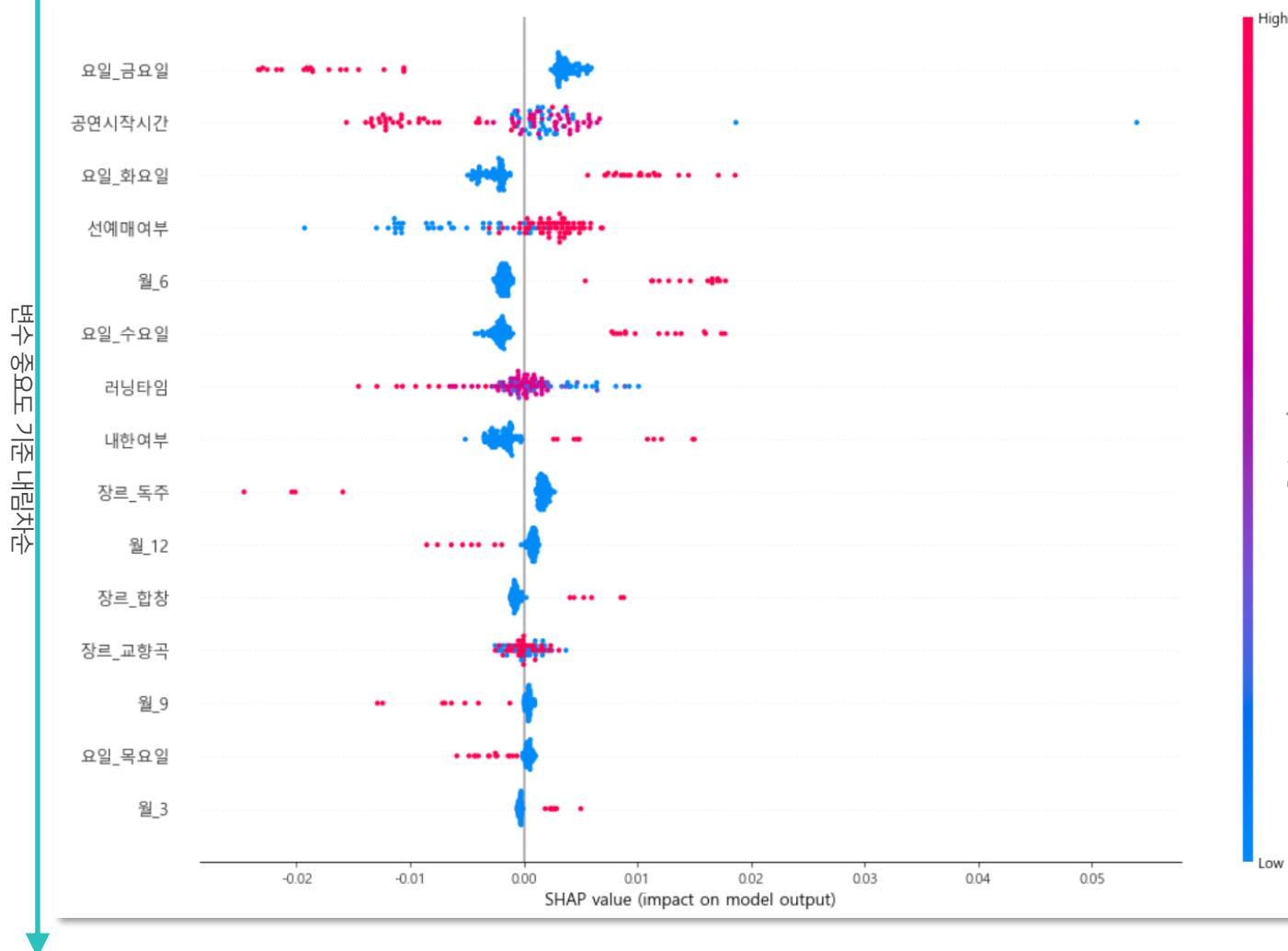
: ML 모델에서 각 변수를 하나씩 빼고 더하면서
해당 변수가 y 예측의 정확도에 얼마나 기여하는지 계산하여
각 변수의 중요도를 수치화하는 방법

중요도 Top 20 변수

	분류	변수
범주형	요일	금요일, 화요일, 수요일, 목요일, 토요일
	월	6월, 12월, 9월, 3월, 8월, 10월, 5월
	장르	독주, 합창, 교향곡, 클래식
	선예매여부	bool
	내한여부	bool
연속형		공연시작시간, 러닝타임



변수 중요도 해석 - (1) R(일반예매율) 예측 : SHAP Summary Plot



SHAP value 해석 방법

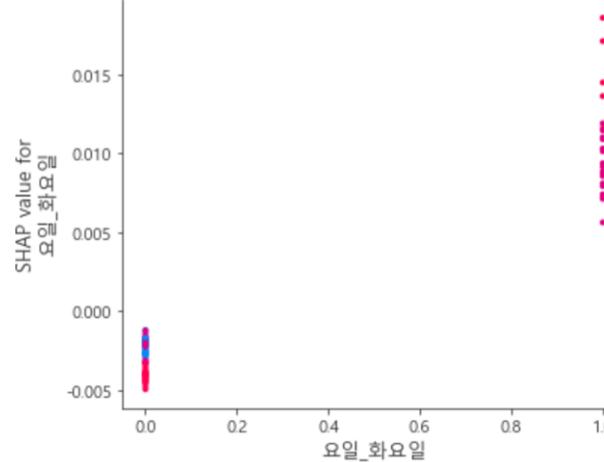
- 색상 : 각 변수의 기여 방향
 - 빨간색 : 변수 값이 높으면
 - 파란색 : 변수 값이 낮으면
- 영역 : y 변수에 대한 작용 방향
 - 양의 영역 : 예측 결과에 긍정적 요인
 - 음의 영역 : 예측 결과에 부정적 요인

※ 단, 인과관계가 아닌 기여 방향의 경향성으로 해석해야 함

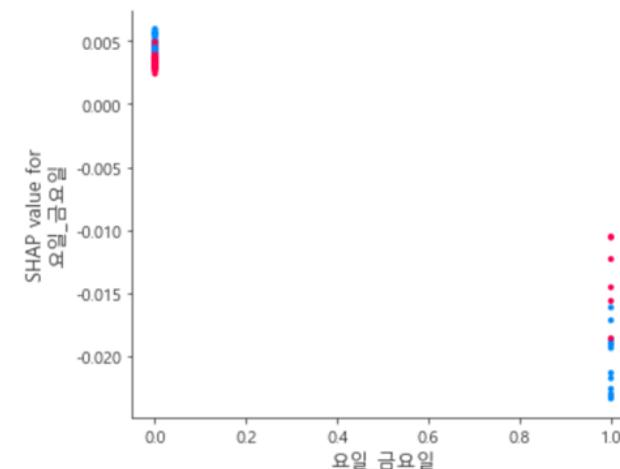
변수별 SHAP value 해석 예시

- 요일_금요일이면 일반예매율이 낮은 경향 있음
- 공연시작시간이 늦으면 일반예매율이 낮은 경향 있음
- 요일_화요일이면 일반예매율이 높은 경향 있음
- 선예매여부이면 일반예매율이 높은 경향 있음
- 월_6이면 일반예매율이 높은 경향 있음
- 요일_수요일이면 일반예매율이 높은 경향 있음

변수 중요도 해석 - (1) R (일반예매율) 예측 : SHAP Dependence Plot



요일_화요일이면
일반예매율이 **높은** 경향 있음
(일반예매율에 긍정적인 영향)



요일_금요일이면
일반예매율이 **낮은** 경향 있음
(일반예매율에 부정적인 영향)

Top15 변수별 SHAP value 해석

예측 기여 경향성	중요도 Top15 변수
값이 커지면 일반예매율이 높은 경향이 있는 변수들	<ul style="list-style-type: none"> 요일_화요일* 선예매여부* 월_6* 요일_수요일* 내한여부* 장르_합창 월_3
값이 커지면 일반예매율이 낮은 경향이 있는 변수들	<ul style="list-style-type: none"> 요일_금요일* 공연시작시간* 러닝타임* 장르_독주* 월_12* 월_9 요일_목요일
해석 불가한 변수들	<ul style="list-style-type: none"> 장르_교향곡

* : 중요도 Top10 변수

변수 중요도 해석 - (2) M (멤버십 유입률) 예측 : SHAP Feature Importance

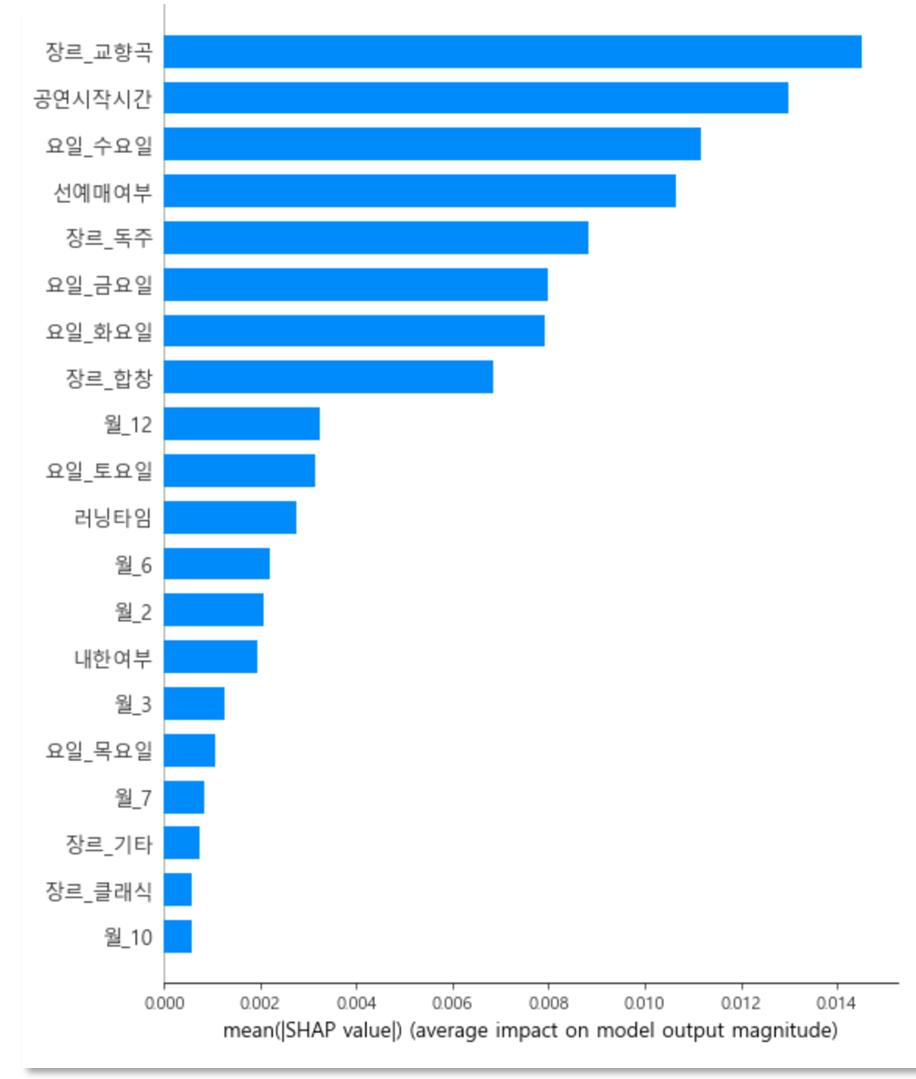
SHAP 변수 중요도

Shap (SHapley Additive exPlanations)

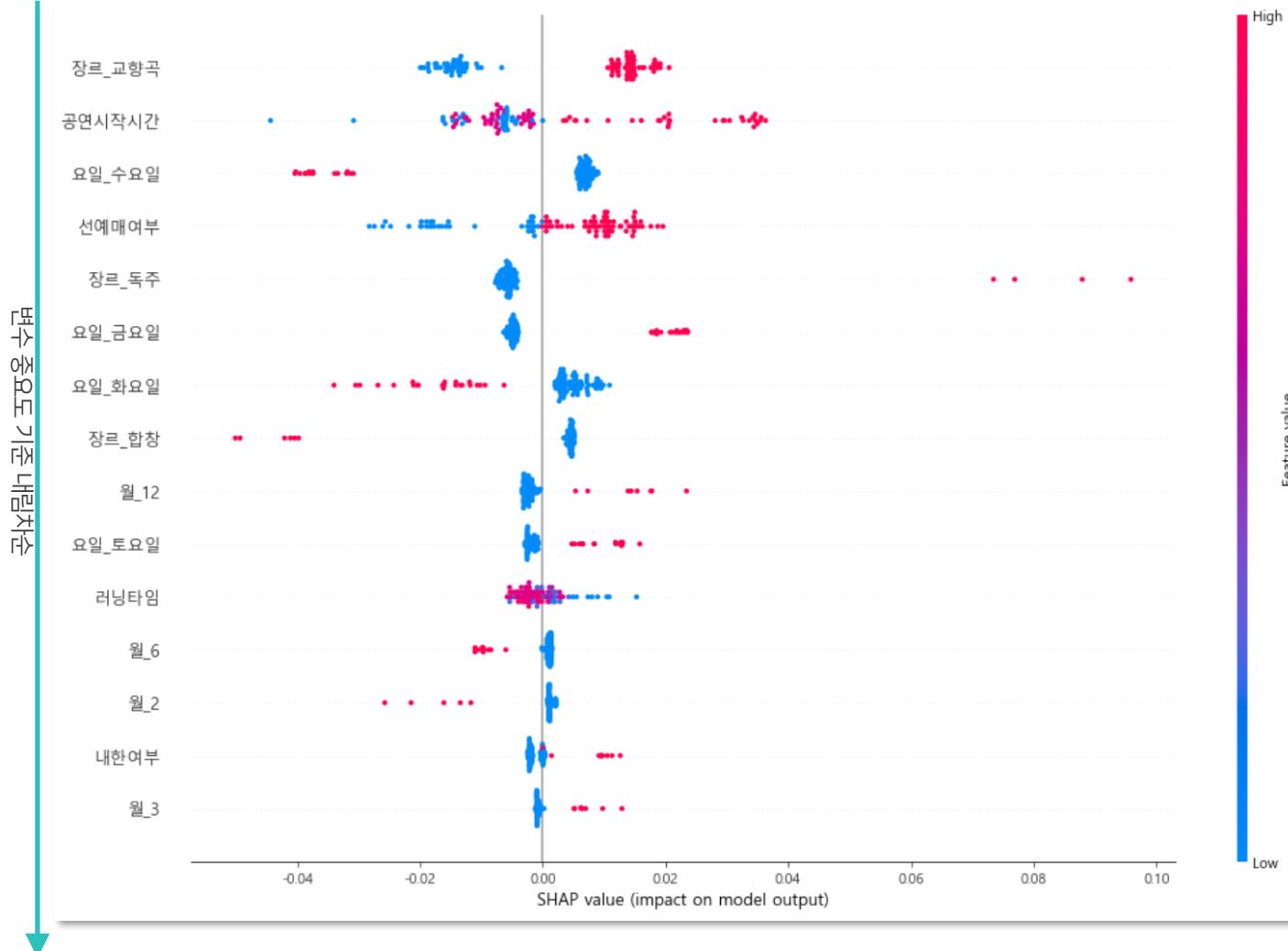
: ML 모델에서 각 변수를 하나씩 빼고 더하면서
해당 변수가 y 예측의 정확도에 얼마나 기여하는지 계산하여
각 변수의 중요도를 수치화하는 방법

중요도 Top 20 변수

	분류	변수
범주형	요일	수요일, 금요일, 화요일, 토요일, 목요일
	월	12월, 6월, 2월, 3월, 7월, 10월
	장르	교향곡, 독주, 합창, 기타, 클래식
	선예매여부	bool
	내한여부	bool
연속형		공연시작시간, 러닝타임



변수 중요도 해석 - (2) M (멤버십 유입률) 예측 : SHAP Summary Plot



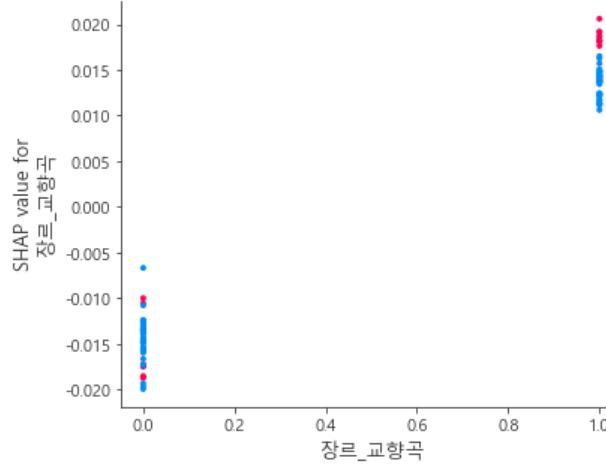
SHAP value 해석 방법

- 색상 : 각 변수의 기여 방향
 - 빨간색 : 변수 값이 높으면
 - 파란색 : 변수 값이 낮으면
 - 영역 : y 변수에 대한 작용 방향
 - 양의 영역 : 예측 결과에 긍정적 요인
 - 음의 영역 : 예측 결과에 부정적 요인
- ※ 단, 인과관계가 아닌 기여 방향의 경향성으로 해석해야 함

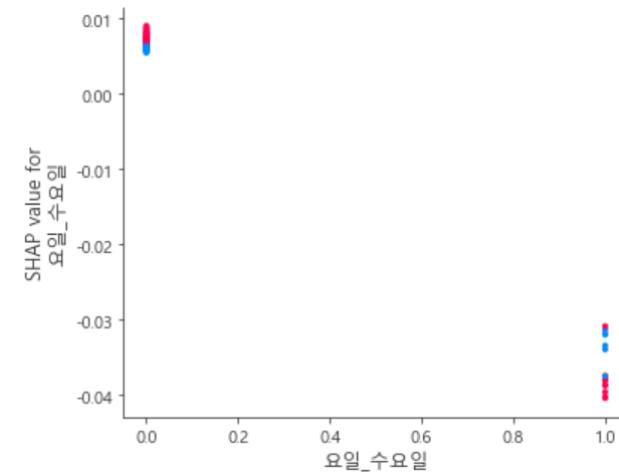
변수별 SHAP value 해석 예시

- 장르_교향곡이면 멤버십 유입률이 **높은 경향** 있음
- 공연시작시간이 **늦으면** 멤버십 유입률이 **높은 경향** 있음
- 요일_수요일이면 멤버십 유입률이 **낮은 경향** 있음
- 선예매여부이면 멤버십 유입률이 **높은 경향** 있음
- 장르_독주이면 멤버십 유입률이 **높은 경향** 있음
- 요일_금요일이면 멤버십 유입률이 **높은 경향** 있음

변수 중요도 해석 - (2) M (멤버십 유입률) 예측 : SHAP Dependence Plot



장르_교향곡이면
멤버십 유입률이 **높은** 경향 있음
(멤버십 유입률에 긍정적인 영향)



요일_수요일이면
멤버십 유입률이 **낮은** 경향 있음
(멤버십 유입률에 부정적인 영향)

Top15 변수별 SHAP value 해석

예측 기여 경향성	중요도 Top15 변수
값이 커지면 멤버십 유입률이 높은 경향이 있는 변수들	<ul style="list-style-type: none"> 장르_교향곡* 공연시작시간* 선예매여부* 장르_독주* 요일_금요일* 월_12* 요일_토요일* 내한여부 월_3
값이 커지면 멤버십 유입률이 낮은 경향이 있는 변수들	<ul style="list-style-type: none"> 요일_수요일* 요일_화요일* 장르_합창* 러닝타임 월_6 월_2

* : 중요도 Top10 변수

Part 6

가격 모델링

- 6.1 가격 모델링 개요
- 6.2 가격 산출 공식
- 6.3 가격 산출 예시



가격 모델링 개요

1

공연 특성에 따른 클러스터링

2

클러스터 내 평균 가격으로 기준 가격 설정

3

예매율과 멤버십 유입률로 기준 가격 보정

4

[최저가~최고가] 범위 내에서 좌석 등급에 따라 가격 차등 분배

가격 산출 공식 - (1) 공연 특성에 따른 클러스터링

공연시작시간

선예매여부

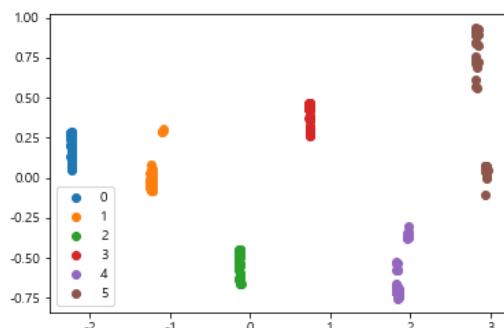
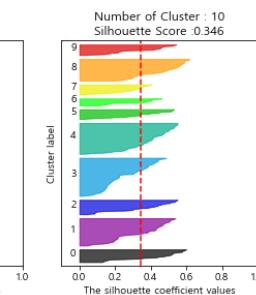
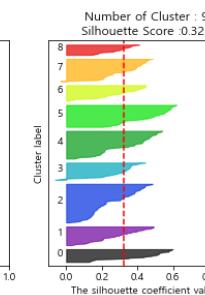
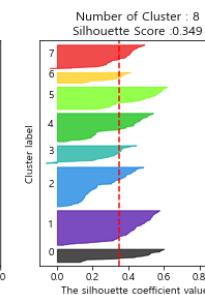
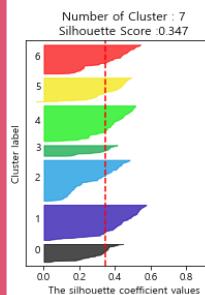
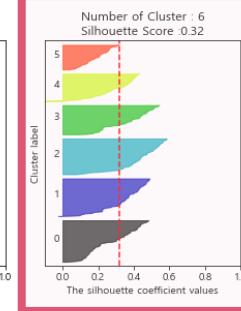
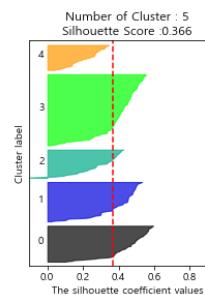
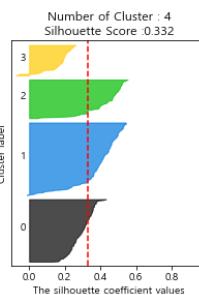
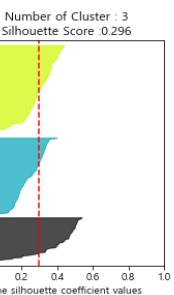
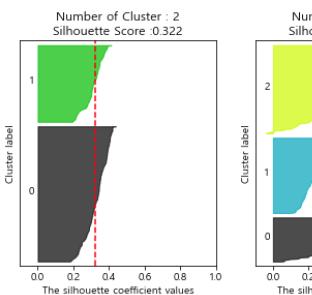
장르

내한여부

러닝타임

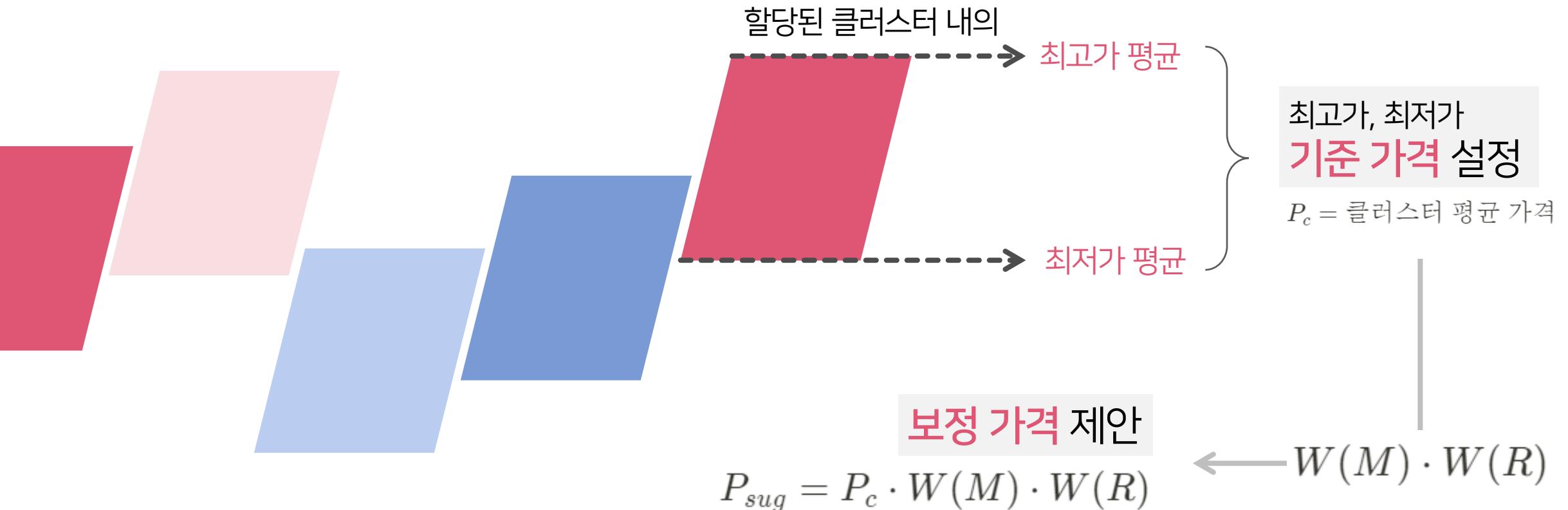
요일

월



K-Means 클러스터링 후 실루엣 계수를 분석하여
총 **6개**의 공연 군집으로 클러스터링

가격 산출 공식 - (2) 클러스터 내 평균 가격으로 기준 가격 설정



가격 산출 공식 - (3) 예매율과 멤버십 유입률로 기준 가격 보정

$$P_{sug} = P_c \cdot W(M) \cdot W(R)$$

가중 함수를 활용하여 기준 가격 보정

“ $W(X) = e^{\hat{X}} - e^{E[X]} + 1$ ”

일반 예매율

$$R = \frac{n}{N}$$

멤버십 가중 예매율

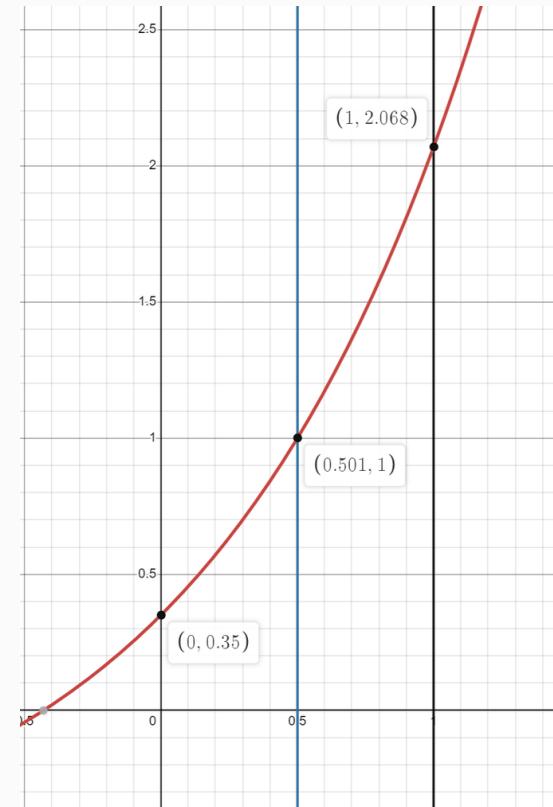
$$R_w = \frac{\sum_{i=1}^n m_i}{25N}$$

멤버십 유입률

$$M = \frac{R_w}{R} = \frac{\sum_{i=1}^n m_i}{25n}$$

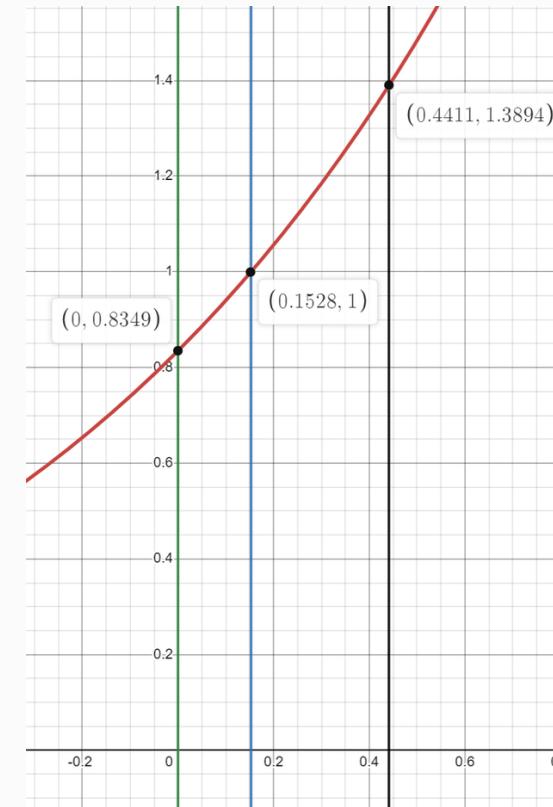
N = 오픈 좌석 수 n = 판매 좌석 수
 m_i = i 번째 예매 데이터의 멤버십 점수

일반예매율 가중 함수



$$(e^{\hat{R}} - e^{E[R]} + 1)$$

멤버십유입률 가중 함수



$$(e^{\hat{M}} - e^{E[M]} + 1)$$

가격 산출 공식 - (4) [최저가 ~ 최고가] 범위 내에서 좌석 등급에 따라 가격 차등 분배

	개수	차이	차이비율	가격차	가격
R	800	200	차이 / 차이 총합	3.6만	10만
A	600	150	0.4 *(최고가 - 최저가)	2.7만	6.4만
B	450	100	0.3	1.8만	3.7만
C	350	50	0.2	0.9만	1.9만
D	300				1만

좌석등급

최고가
- 3.6만
- 2.7만
- 1.8만
- 0.9만
최저가

R 개수 - A 개수
A 개수 - B 개수
B 개수 - C 개수
C 개수 - D 개수

차이 / 차이 총합
차이비율 *(최고가 - 최저가)

가격 산출 예시

공연 일자	공연 시작 시간	선 예 매 여 부	예매시작 일	장 르	휴 게 시 간	공연명	내 한 여 부	러 닝 타 임	가 격 종 류	R석	A석	요 일	일반예 매율	멤버십 유입률	월	cluster_R석	cluster_A석	P _c	P _{sug}
										예술의 전당 전 관 개관 30주년 특별음 악회-코 리안챔 버오케 스트라 초청 콘 서트	목 요 일	0.51146	0.12985	10	81696.96970	25454.54545	80940.83014	25218.95301	
1 2023-10-19	1930	1	20230823	클래식	0	악회-코리안챔버오케스트라초청 콘서트	0	120	4	80000	20000	목 요 일	0.51146	0.12985	10	81696.96970	25454.54545	80940.83014	25218.95301
2 2023-11-04	1700	0	20230901	독주	20	바리톤 김기훈 리사이틀	0	90	3	100000	60000	토 요 일	0.47639	0.22579	11	89978.72340	26936.17021	94016.35509	28144.88190
4 2023-12-01	2000	1	20221210	교향곡	15	2023 서울시향 압판 츠베텐의 베토벤 삼중 협주곡 ②	1	100	5	100000	10000	금 요 일	0.45764	0.26033	12	150553.84615	43876.92308	158579.28020	46215.82946

가격 산출 예시 - (1) 2023.10.19 클래식 공연

① 가격을 산출하고자 하는 공연 정보 수집

변수	값
공연일자	2023-10-19(목)
공연시작시간	1930
선예매여부	1
예매시작일	20230823
장르	클래식
휴게시간	0
공연명	예술의전당 전관 개관 30주년 특별음악회-코리안챔버오케스트라 초청 콘서트
내한여부	0
러닝타임	2시간

② 예측 R(일반예매율), M(멤버십유입률) & 할당 클러스터 기준가

R(일반예매율)	M(멤버십유입률)	cluster_R석	cluster_A석
0.51146	0.12985	81696.96970	25454.54545

③ 최고가 · 최저가 계산 & 장르별 좌석 등급별 가격 차등 분배

변수	값
제안가격_최고가	80940.83014
제안가격_최저가	25218.95301

클래식 좌석	R석	A석	B석	C석	D석
개수	194	234	490	674	915

∴ 클래식 장르의 좌석 그룹핑 : 5단계



좌석 등급	가격(원)	개수(개)
R석	80,941	194
A석	77,850	234
B석	58,065	490
C석	43,845	674
D석	25,219	915

가격 산출 예시 - (2) 2023.11.04 독주 공연

① 가격을 산출하고자 하는 공연 정보 수집

변수	값
공연일자	2023-11-04(토)
공연시작시간	1700
선예매여부	0
예매시작일	20230901
장르	독주
휴게시간	20
공연명	바리톤 김기훈 리사이틀
내한여부	0
러닝타임	1시간 30분

② 예측 R(일반예매율), M(멤버십유입률) & 할당 클러스터 기준가

R(일반예매율)	M(멤버십유입률)	cluster_R석	cluster_A석
0.47639	0.22579	89978.72340	26936.17021

③ 최고가 · 최저가 계산 & 장르별 좌석 등급별 가격 차등 분배

변수	값
제안가격_최고가	94016.35509
제안가격_최저가	28144.88190

독주 좌석	R석	A석	B석	C석	D석
개수	373	557	654	441	478

:: 독주 장르의 좌석 그룹핑 : 5단계



좌석 등급	가격(원)	개수(개)
R석	94,016	373
A석	71,191	557
B석	59,158	654
C석	32,735	441
D석	28,145	478

가격 산출 예시 - (3) 2023.12.01 교향곡 공연

① 가격을 산출하고자 하는 공연 정보 수집

변수	값
공연일자	2023-12-01(금)
공연시작시간	2000
선예매여부	1
예매시작일	20221210
장르	교향곡
휴게시간	15
공연명	2023 서울시향 압판 츠베덴의 베토벤 삼중 협주곡 ②
내한여부	1
러닝타임	1시간 40분

② 예측 R(일반예매율), M(멤버십유입률) & 할당 클러스터 기준가

R(일반예매율)	M(멤버십유입률)	cluster_R석	cluster_A석
0.45764	0.26033	150553.84615	43876.92308

③ 최고가 · 최저가 계산 & 장르별 좌석 등급별 가격 차등 분배

변수	값
제안가격_최고가	158579.28020
제안가격_최저가	46215.82946

교향곡 좌석	R석	A석	B석	C석	D석
개수	194	221	560	673	861

∴ 교향곡 장르의 좌석 그룹핑 : 5단계



좌석 등급	가격(원)	개수(개)
R석	158,579	194
A석	154,031	221
B석	96,923	560
C석	77,887	673
D석	46,216	861

Part 7

결론

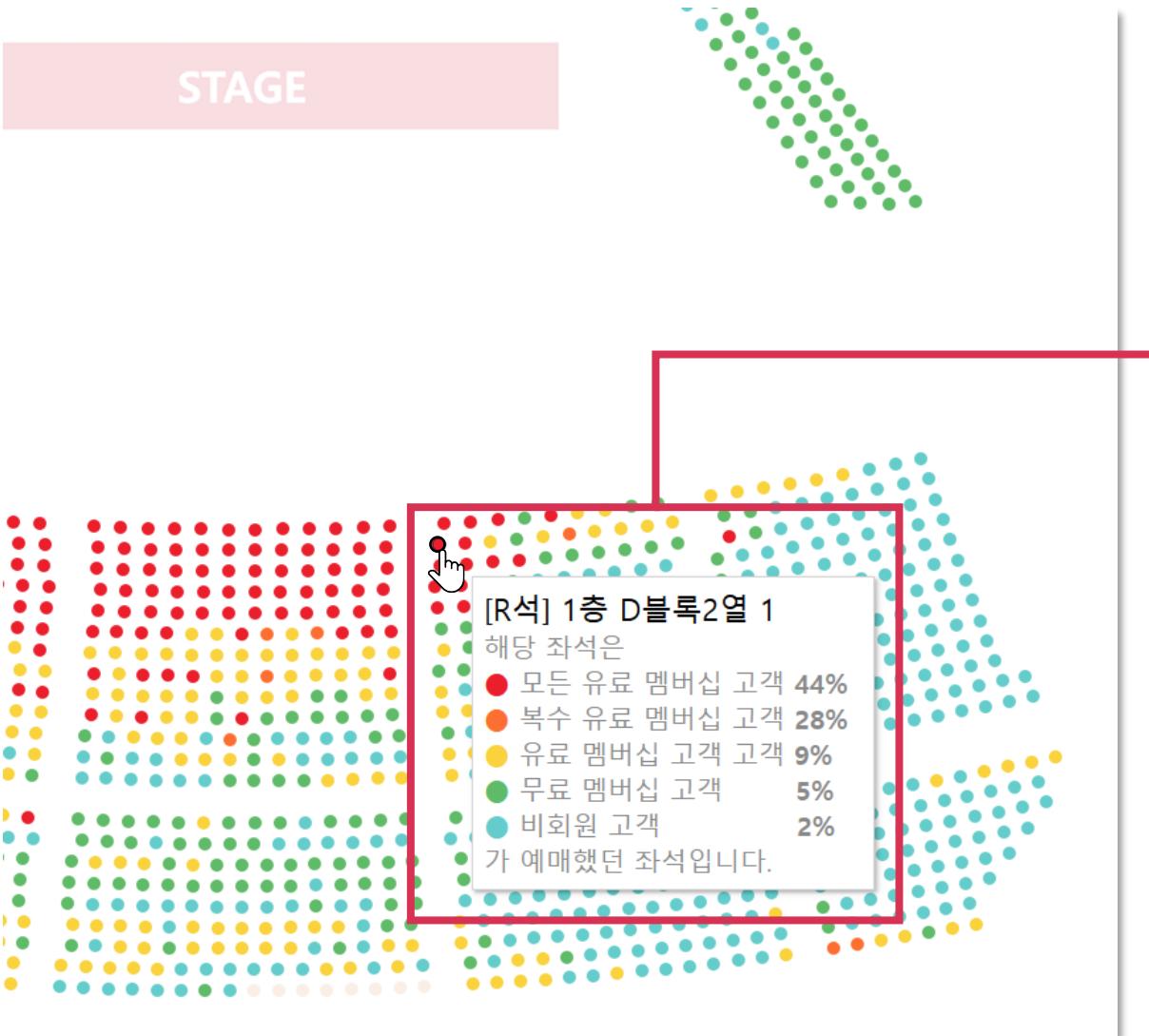
7.1 활용방안 & 기대효과

7.2 의의 및 한계



활용방안

STAGE



새로운 공연 정보 → 좌석 그룹핑과 가격 산출 자동화

좌석을 선택했을 때
좌석 등급과 함께
'어떤 고객이 얼마나 예매했었는지'와 같이
좌석에 대한 추가적인 데이터 제공

가격 설정 근거 제공 → 신뢰성 확보
추가적인 데이터 → 구매 결정에 도움

주체별 기대효과

- 예측 수요에 따라 가격 보정을 함으로써
수요와 수익을 최대화하는
최적의 가격 설정 가능

- 합리적인 가격 및 좌석 구분을 통한
고객 만족도 향상, 공연장 재방문 유도
및 클래식 공연 활성화

- 클래식 공연에 대한 관심도에 따라
적절한 좌석을 선택할 수 있는
가이드라인 제시

- 개개인의 지불 용의 내에서
최대의 소비 효용을 달성할 수 있음

예술의 전당

관람객

의의 및 한계

의의

- 일반예매율과 더불어 멤버십을 고려한 유입률 등 **다양한 파생변수**를 가격 모델에 활용
- 태블로에서의 좌석 시각화를 통해 직관적으로 다양한 요소에 따른 좌석의 위치별 선호도 파악
- **멤버십**을 활용하여 공연에 대한 **관심도가 다양한 관람객**에 맞춤형으로 대응할 수 있는 가격 제안

한계

- 출연진, 기획사의 인지도를 파악할 수 있는 데이터의 부재
- 원본 데이터의 가명정보처리로 인한 매칭 오차 발생

Part 8

툴·데이터 및 참고문헌



활용 툴·데이터 및 참고문헌 출처



활용 툴	활용 분야
Tableau	데이터 시각화
Python (pandas, numpy, matplotlib, seaborn, scipy, scikit-learn, seaborn, skimage, shap 등)	데이터 시각화, 분석, 모델링

데이터	제공 플랫폼	데이터 상세
KOPIS API	KOPIS, 공연예술통합전산망	공연목록, 공연상세, 일별 티켓판매수 및 티켓판매액
예술의전당 공연일정	예술의전당	장르, 공연명, 기간, 장소

기사/칼럼	기사 상세
"30만~40만원으로 껑충" 비싸진 클래식 공연티켓값, 원인은 '항공 운임'	https://www.asiae.co.kr/article/2022070422072908985
日보다 70% 비싼 韓클래식 티켓값…“문화강국 되려면 문턱 낮춰야”	https://www.hankyung.com/life/article/2022070610091
음악회의 좋은 자리를 찾는 꿀팁!	https://brunch.co.kr/@truth-art/24
클래식 음악회 공연장별 꿀좌석은 어디?	https://www.khan.co.kr/culture/culture-general/article/201908301402011
[칼럼] 국내 주요 클래식 콘서트홀 리뷰1 - 예술의전당 콘서트홀	https://www.artinsight.co.kr/news/view.php?no=65604

논문	
오경지. (2013). <클래식 음악공연의 좌석등급 및 티켓 가격 인플레 문제 연구>. 이화여자대학교공연예술대학원 석사학위논문.	

Ye, Peng, et al. "Customized Regression Model for Airbnb Dynamic Pricing." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, 2018, pp. 932–940.

보고서 등 기타 자료	자료 상세
Airbnb Dynamic Pricing Optimization	https://github.com/tule2236/Airbnb-Dynamic-Pricing-Optimization/tree/master
7 Common Pricing Models	https://www.indeed.com/career-advice/career-development/pricing-modeling
Dynamic Pricing using ML	https://medium.com/@baabak/dynamic-pricing-using-machine-learning-5e882282effe
Implementing Dynamic pricing strategy in python	https://induraj2020.medium.com/implementing-dynamic-pricing-strategy-in-python-part-1-5bd7e1a1f382
Ticket Sales by Seat Type Report for Sports Organizations and Venues	https://www.solverglobal.com/report-budget-forecast-and-dashboard-template-glossary/ticket-sales-by-seat-type-report-for-sports-organizations-and-venues/
Analysing Cinema Seating Patterns	https://medium.com/analytics-vidhya/popcorn-data-analysing-cinema-seating-patterns-part-ii-987fbde9d363
하우스 매니저가 알려주는 음악당 객석 꿀팁	https://www.youtube.com/watch?v=k0kALhTkbKo